



TAMPERE UNIVERSITY OF TECHNOLOGY

**Kamiar Radnosrati**  
**Trading-off Compression, Energy Efficiency, and QoE In**  
**Wireless Network**

Master's Thesis

Examiner: Prof. Yevgeni Koucheryavy  
Dr. Dmitri Moltchanov  
Examiner and topic approved by the  
Faculty Council of the Faculty of  
Computing and Electrical  
Engineering On March 6, 2013

# ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY

Master's Degree Programme in Information Technology

**Kamiar Radnosrati : Trading-off Compression, Energy Efficiency, and QoE In Wireless Network**

Master of Science Thesis, 71 pages

April 2013

Major: Communications Engineering

Supervisor: Dr. Dmitri Moltchanov

Keywords: Energy-aware communication, QoE, Media applications, Energy efficiency, Complexity

While throughput of wireless networks is improving rapidly during recent years, handheld devices are becoming more and more powerful and small. These features made multimedia streaming a favorite application of mobile smart devices. However, their lifetime highly depends on their battery powers which is not growing at the same pace of communication technologies. That is why, finding a solution to optimize energy consumption while keeping quality at the best possible level in multimedia applications is of high importance.

This work qualitatively and quantitatively describes the trade-off between the total energy required for communication (operation of the service, encoding and transmission), compression ratio of different audio and video codecs, and obtained quality from the network in wireless environment. The main aim of this project is to investigate the potential of implementing a system that minimizes power consumption while maintaining the best possible QoE.

Possibility of designing a system that is capable of adjusting its parameters based on the battery level of device and the desired QoE would also be examined. Appropriate modification of parameters in multimedia application management system enables service providers to choose the most optimized algorithm in different conditions.

Results of the project will provide unified solution for joint performance and energy optimization for media applications running over wireless channels. A real-time system responsible for dynamic optimization of energy and QoE metrics can be developed in future studies.

## PREFACE

This project is the final proof of Master of Science (MSc) Thesis, Trading-off Compression, Energy Efficiency, and QoE In Wireless Network, in the Department of Communications Engineering at Tampere University of Technology, Finland. The executed research has been done during the year 2012-2013 at the Institute of Communication Engineering, Tampere University of Technology.

Several people have made this possible being supportive of me. I would like to thank my supervisor, Dr. Dmitri Moltchanov for his time, valuable guidance, and constructive feedback. Thanks to my friends at Tampere for sharing their experience and being helpful during the whole process of this project. Without them it was very difficult handling obstacles occurred during this time.

Last, but definitely not least, I would like to express my gratitude to my family for motivating and supporting me in every possible way they could.

Oulu, April 2013

Kamiar Radnosrati  
kamiar.radnosrati@tut.fi  
Yliopistokatu 28 C 117,  
90570, Oulu,  
Finland  
Tel +358 46 526 8162

# CONTENT

1. Introduction . . . . .	1
1.1 Motivation . . . . .	2
1.2 Energy-efficient communication . . . . .	3
1.3 Modern media applications and QoE . . . . .	4
1.4 Structure of the thesis . . . . .	7
2. Description of Trade-offs . . . . .	8
2.1 Media Compression . . . . .	8
2.2 Quality of User Experience . . . . .	9
2.2.1 Speech compression . . . . .	9
2.2.2 Video Compression . . . . .	11
2.2.3 Encoding QoE Degradation . . . . .	13
2.2.4 Network QoE Degradation . . . . .	14
2.3 Energy Requirements . . . . .	16
2.3.1 Energy for encoding . . . . .	16
2.3.2 Energy for transmission . . . . .	17
2.3.3 Energy for different technologies . . . . .	19
3. Audio Codecs . . . . .	21
3.1 VoIP challenges and protocols . . . . .	21
3.1.1 QoS weak-points . . . . .	22
3.1.2 Standards and protocols . . . . .	24
3.2 Codecs compression technologies and their applications . . . . .	26
3.3 Trade-offs in audio codecs . . . . .	30
3.3.1 Compression . . . . .	30
3.3.2 QoE . . . . .	33
3.3.3 Energy Consumption . . . . .	35
4. Video Codecs . . . . .	40
4.1 Video communications . . . . .	40
4.1.1 Monitoring quality of communication . . . . .	40
4.1.2 Compression challenges . . . . .	41
4.2 H.264 . . . . .	43
4.2.1 Codec Structure . . . . .	44
4.2.2 Profiles and levels . . . . .	47
4.3 Trade-offs in H.264 . . . . .	49
4.3.1 Methodology . . . . .	49
4.3.2 Energy required for compression and transmission . . . . .	51
4.3.3 PSNR and SSIM . . . . .	53
5. Conclusion . . . . .	58

## ABBREVIATIONS AND DEFINITIONS

QoE	Quality of Experience
WSN	Wireless Sensor Network
4G	4th Generation
IP	Internet Protocol
ARQ	Automatic Repeat-reQuest
FEC	Forward Error Correction
PDU	Protocol Data Unit
MIMO	Multiple-Input and Multiple-Output
SNR	Signal to Noise Ratio
ICT	Information and Communications Technology
QoS	Quality of service
ITU	International Telecommunication Union
VoIP	Voice over IP
RLE	Run Length Encoding
VQ	Vector Quantization
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
PSNR	Peak Signal to Noise Ratio
SSIM	Structural SIMilarity
MOS	Mean Opinion Score
PLR	Packet Loss Ratio
TVI	Time Varying Impairments
RE	Recency Effect
VQA	Video Quality Assessment

MSE	Mean Square Error
HVS	Human Visual System
MIPS	Million Instructions Per Second
EC	Error Concealment
DSP	Digital Signal Processor
CS-ACELP	Conjugate Structure Algebraic Code-Excited Linear Prediction
ADPCM	Adaptive Differential Pulse Code Modulation
QCIF	Quarter Common Intermediate Format
SISO	Single Input Single Output
LTE	Long-Term Evolution
TTI	Transmission Time Interval
UMTS	Universal Mobile Telecommunication System
PSTN	Public Switched Telephone Network
SIP	Session Initiation Protocol
BER	Bit Error Rate
DoS	Denial of Service
IETF	Internet Engineering Task Force
SDP	Session Description Protocol
TCP	Transmission Control Protocol
PPS	Packet Per Second
MDCT	Modified Discrete Cosine Transform
BWE	BandWidth Extension
AVQ	Algebraic Vector Quantization
FLVQ	Fuzzy Learning Vector Quantization
TD	Time Differential

FD	Frequency Differential
MLT	Modulated Lapped Transform
TV	Time Variant
cRTP	compressed Real-time Transport Protocol
MLP	Multilink Point-to-Point Protocol
FRF	Frame Relay Forum
FCS	Frame Check Sequence
CRC	Cyclic Redundancy Check
PSQM	Perceptual Speech Quality Measure
PESQ	Perceptual Evaluation of Speech Quality
PAMS	Perceptual Analysis Measurement System
MOPS	Million Operations Per Second
WMOPS	Weighted Million Operations Per Second
ETSI	European Telecommunications Standardization Institute
OFDMA	Orthogonal Frequency Division Multiple Access
W-CDMA	Wide-band Code Division Multiple Access
RTP	Real-time Transport Protocol
RTCP	RTP Control Protocol
ASF	Advanced Systems Format
AVI	Audio Video Interleave
ISO	International Standards Organization
CAVLC	Context-Adaptive Variable-Length Coding
CABAC	Context-Adaptive Binary Arithmetic Coding
SVC	Scalable Video Coding
MVC	Multi-view Video Coding

SBP	Scalable Baseline Profile
SHP	Scalable High Profile
CP	Complexity Parameter
QP	Quantization Parameter



# 1. INTRODUCTION

VoIP and video streaming, including video telephony, currently dominate the market. Because these applications have strict requirements regarding delay and available bandwidth, any changes in the network might degrade user satisfaction and Quality of Experience (QoE). On the other hand, a considerable number of these applications run over hand-held devices through WiFi or cellular networks. These devices are limited in their energy resources and, as a result, suffer from energy consumption limitations. This issue forces manufacturers to use processors supporting codecs, as these processors consume less energy.

Each codec is characterized by a specific set of properties. With some codecs, the energy consumed for encoding/decoding, called complexity of codec, is reasonable, while the output quality may be inadequate or even not applicable in IP network applications. Improving the quality might require providers to use more complex compression algorithms. The resulting higher complexity level, however, could increase energy consumption. Given these considerations, a comprehensive study examining these features of the codec specifications would help to determine the most efficient QoE spending and energy consumption.

In this chapter, we emphasize the need for energy efficient communication in energy constrained environments. Different environments might require energy consumption optimization in order to operate more efficiently. Wireless Sensor Network (WSN) present a field in which different studies regarding energy efficiency have been performed.

Statistics of daily usage of the Internet prove that WSN is not the only field in which energy consumption must be investigated accurately. One of the most important areas consuming noticeable energy in different phases of operation is multimedia over wireless in all wireless-access-ready devices, specifically portable, rechargeable devices. This would include laptops, smart phones, and tablets, which provide media on demand and different conferencing environments.

We discuss the potential of trade-offs between network-related energy expenses and application related ones, as each of these imposes its own demands on the total energy consumption. Considering these expenses independently would not result in an optimized solution for total energy expenses reduction.

In the context of network-related issues, transmission technologies, on, off, and

idle periods of transmission, the specific energy consumption model in each scenario, and the duration of each of these cycles are important. With application-related issues, different algorithms used for digitizing analogue signals and the rate at which this digitization is done must be examined carefully.

Application-related energy expenses stem from compression and coding required by codecs, to process pre-defined algorithms. In some applications, these methods are defined and changing them is not possible; in others, there is the possibility of changing between different codecs that may have different specifications and use different compression techniques. Sometimes it is also possible to modify an existing codec to obtain better performance in terms of quality and energy efficiency.

## 1.1 Motivation

With the commercial launch of 4th Generation (4G) mobile systems, the convergence of the Internet and mobile communications is becoming clear. This process is stimulated by the rapid growth of both technologies, as well as by users who wish to readily access Internet services. In addition to providing broadband wireless access to the Internet, modern and future wireless networks are expected to satisfy QoE requirements of various media applications. This is an inherent problem for many service types, even in fixed Internet Protocol (IP) networks. Wireless and mobility issues add their own problems on top of this IP flaw.

The time-varying erroneous nature of wireless channels, cross-layer organization of the protocol stack, and tele-traffic and mobility issues should be addressed before wireless media Internet services can be widely deployed. The intention to adopt IP protocol for future mobile communications and the extension of Internet services to the air interface necessitates novel performance evaluation and optimization methods.

Since the error rate of a fixed transmission medium is negligibly low, to evaluate the performance of applications in fixed IP networks, it is sufficient to estimate performance degradation caused by packet forwarding procedures. When dealing with wireless networks, we must also take into account performance degradation caused by bit errors of wireless channels. This degradation, which significantly contributes to end-to-end performance expectations, has a completely different nature than that seen in fixed networks. The air interface is expected to be a weak point in any QoE assurance model.

To provide optimized performance, wireless technologies implement a number of channel adaptation mechanisms. These are Automatic Repeat-reQuest (ARQ) and Forward Error Correction (FEC) techniques, adaptive size of Protocol Data Unit (PDU)s at different layers, Multiple-Input and Multiple-Output (MIMO) antenna design, power control mechanisms, etc. All of these mechanisms are implemented at

different layers of the protocol stack and affect performance provided to applications differently. To provide their unified and optimized operation, wireless technologies call for novel design of the protocol stack that should include cross-layer performance optimization capabilities.

The recent improvement in processing and storage capabilities of mobile devices makes them suitable for computationally media intensive services, e.g., video streaming. Such services are characterized by strict QoE metrics, but add another complicating factor related to excessive energy consumption of mobile devices. The energy should be spent not only on transmission of actual data, but for encoding/decoding operations. Given the dearth of breakthroughs, in battery technology, this problem will only get worse as more capable devices and services hit the market. What is more important is that the problem of providing satisfactory QoE is inherently related to the energy consumption of various transmission technologies, and this relationship is not linear.

Theoretical and practical results of the project will provide basic ideas of a unified solution for joint performance and energy optimization for media applications running over wireless channels. A real-time system responsible for dynamic optimization of energy and QoE metrics can be developed. The performance control and optimization system, if properly formulated, designed, and implemented, can be of high practical importance.

Focusing on energy considerations makes it possible to design a potential system that may trade-off energy for air interface with one factor used for compression and coding. A QoE-ware system should use the most appropriate codec (compression algorithm) to optimize power consumption and ensure a desired level of quality.

This total energy is the addition of transmission and compression and coding. Each coding method results in a certain level of QoE in terms of a specific level of energy. Defining a threshold for both QoE and energy expenses allows the potential system to modify its transmission technology or complexity of codec in order to make the best choice that satisfies all limitations.

## 1.2 Energy-efficient communication

In the context of energy consumption in wireless networks, the effect of communication equipment cannot be neglected as a major energy consuming component [1]. Two main factors to be modified in order to reduce energy consumption are transmission time and transmission power. Increasing transmission bitrate results in lower transmission time but requires more transmission power to satisfy Signal to Noise Ratio (SNR) constraints. This means that transmission time and power are related, and trade-offs between these two must be examined. [2] investigated both theoretical formulas given by Shannon and their own method for energy consump-

tion in wireless networks.

Considering the previously mentioned power consumption issues, there are two major solutions to remedy the short lifetime of portable devices. The first is to improve the design and capacity of batteries. However, these devices are small and lightweight; as a result, designing batteries is a challenging task.

It is important to mention that the energy density of batteries has improved in recent years, but not at the same pace of Information and Communications Technology (ICT) devices and wireless revolutions [3], [4]. In fact, new batteries are mainly used on higher power consuming devices to maintain runtime of device at the same level of old, less energy consuming devices, at the best case scenario. This deficiency in their capability has made it necessary for researchers to find and introduce energy efficient algorithms in wireless networks; this is the second solution for increasing the lifetime of new powerful portable devices.

Various protocol optimization schemes might enable power management techniques during transmission. The method introduced in [5] focuses on data link layer and discusses retransmissions in an energy constraint environment. [6] focused on multimedia streaming and the effect of traffic structure within network and radio access technology used for transmission.

In addition to improving the lifetime of portable devices and their performance, there are other important factors motivating researchers to work on energy efficiency in communication networks. One such factor concerns global environment issues. According to [7], 15-20 percent of the entire communications industry and 0.3-0.4 percent of annual carbon dioxide emissions relate to mobile communications.

Although there have been some studies done in this field, none have compared all three metrics of QoE, compression, and energy efficiency. These are the areas addressed in this thesis.

### 1.3 Modern media applications and QoE

For a long time, Quality of Service (QoS) was the basic concept used by researchers to show the performance of applications. However, it was necessary to use a metric to provide a better understanding of a user's experience. QoE, the currently used metric, shows user requirements and level of satisfaction with a specific application.

The concept of QoS defined by International Telecommunication Union (ITU) is the totality of characteristics of a telecommunications service that bear on its ability to satisfy the stated and implied needs of the user of the service (ITU-T Rec. E.800 (2008)). On the other hand, QoE can be defined as how satisfied a user is after using an application.

QoE evaluations are conducted using two different approaches. The first approach, called subjective testing, is based on user perceptions and is usually used in

laboratories. The second approach, which is more objective, is based on instrumental calculations such as E-model in speech communication applications. Subjective tests are time-consuming and expensive; thus, objective tests are favorite solutions to calculate QoE value of an application or service.

In multimedia applications, one of the most important factor affecting QoE is the compression algorithm used in codecs for digitizing and compressing audio and video. Codecs used in multimedia applications mostly use lossy compression techniques; however, the way in which they treat signals can vary. More advanced codecs use better compression schemes because of their advanced methods resulting in higher complexity.

Complexity of codecs is one of the most important factors to consider when choosing a codec for a specific application. Higher complexity (more advanced codecs) might result in a better QoE and compression ratio, which means less energy is required for transmission. However, the increasing complexity of codecs means more energy is required for processing a compression algorithm.

In addition to processing power, another effect of complex codecs that use advanced compression techniques is related to the effect of packet losses on transmitted data. In applications that use these codecs, a huge amount of data would be compressed into packets and be transmitted over the channel. In this scenario, losing a single packet causes more degradation in quality than in codecs with lower compression ratios. In addition to this degradation, losing some consecutive packets might even cause the application to drop the session.

For applications in which video data must be compressed and transmitted, more complex techniques result in a noticeable amount of power consumption because of the inherent characteristics of video contents. However, this issue also exists in the case of speech codecs used in Voice over IP (VoIP) applications, where a more complex codec would require much more energy than simpler codecs.

Describing video compression schemes in details requires dense mathematical equations. However, in order to have a better understanding of why complexity in video codecs is significant, one must understand different technologies and compression algorithms:

1. The most primitive algorithm, Run Length Encoding (RLE), is used for encoding consecutive pixels of the same color. The encoded output would be a single codeword. However, RLE cannot be used widely, as it is not common to have runs of the same color. An example of RLE is encoding the uncompressed sequence "AAAABBBCCDDDDDD," which requires 13 Bytes of storage space; after encoding by RLE, it becomes "4A3B2C4D" (4 of As, 3 of Bs, 2 of Cs, and 6 of Ds). In this example, the compression ratio would be almost 2 to 1.

2. The basic idea of Vector Quantization (VQ) to divide an image into blocks. In this algorithm, different blocks of an image are compared; similar blocks are gathered into one class. Each of these classes is presented using a code in binary. Then, the encoder simply inputs one generic block instead of all of the original blocks of each class. It is then only necessary to encode the binary code of each generic block using the lookup table that maps these blocks to their corresponding binary code.

In order to improve the efficiency of codec, more frequently occurring generic blocks are presented using shorter binary codes. At the other side, the decoder uses the lookup table to reassemble generic blocks from the binary codes. This algorithm is a quick, lossy compression algorithm.

3. Discrete Cosine Transform (DCT) is widely used in different codecs such as MPEG-4 and H.263 video conferencing standard. In this technique, a two-dimensional DCT is applied to  $8 \times 8$  blocks of pixels. Then the resulting coefficients are quantized by a quantization factor. In the next step, RLE is applied to these quantized coefficients. Most DCT coefficients are small; they become zero after the quantization process. Such factors are considered because the human eye is less sensitive to higher frequencies; thus a large quantization factor is applied to higher frequency components.
4. The concept Discrete Wavelet Transform (DWT) involves passing the original signal through a low pass filter to obtain lowered resolution of the signal. The original signal is also passed through a high pass filter to obtain a detailed version of the signal. Using downsampling by two the addition of these two signals does not affect the original number of bits. Choosing appropriate parameters for filters, upsampled versions of two signals results in the original signal. Up to this step, there is no compression, as the original signal with the same number of bits has been regenerated. In order for compression to happen, VQ is applied to the detailed signal coefficients. After this quantization, some coding techniques are also applied to the quantized coefficients; one of these techniques is entropy coding. Based on the quantization factor and the coding algorithm applied to quantized coefficients, the compression ratio can be managed.
5. Motion Compensation is used for coding motions within a scene such as a moving train. In fact, this algorithm is a combination of a number of different algorithms. For compensating purposes, the method chooses a reference picture; this might be a previous or future picture. The method then describes an image based on that reference picture and the current one. In other words,

it uses the fact that in many frames of a movie, most of the objects remain stationary. As a result, it is only necessary to store the information needed to present these transformations.

All of these techniques and those used in speech codecs require different levels of complexity and amounts of energy. They might improve QoE after compression because of more advanced techniques; at the same time, such methods might degrade QoE after transmission due to lost or compressed information. Higher complexity, which mostly means higher compression ratios, requires less transmission energy, but more processing power. That is why choosing the best possible codec depends on many factors, such as network condition, application requirements, and battery capabilities; it is not a straightforward task.

## 1.4 Structure of the thesis

This thesis is organized as follows:

Chapter 2, explains each of three components that this report presents in detail. The first section of Chapter Two describes the reason for compression. Section two describes QoE metrics in both speech and video codecs, and the sources of degradation of QoE are introduced and examined. The last section presents all definitions of energy requirements and compares various compression and transmission power consumption schemes.

Chapter 3, on speech codecs, begins with existing challenges in VoIP. Then different standards and protocols used in VoIP applications are introduced to provide a better overview of these services. Different codecs are compared in terms of the compression algorithms used resulting in various QoE values, processing power, and processing power consumption. Finally, the energy required for transmission over different radio access technologies for each codec is presented and compared to corresponding processing power. The terms "audio" and "speech" are used interchangeably in this thesis.

Chapter 4, video codecs, defines the basic requirements and necessities for communications containing video. The widely used video codec H.264 and the structure of its encoder and decoder are then described briefly. Different profiles and levels of H.264 are suitable for various applications; those most suitable for real-time applications are discussed. Finally, the proposed method for trade-offs in H.264 and the resulting graphs and tables are presented.

Chapter 5, the conclusion, proves that there is a potential for trading-off QoE, compression, and energy consumption that is required when implementing a system that can choose the best codec in different conditions.

## 2. DESCRIPTION OF TRADE-OFFS

In applications such as media-on-demand and multimedia conferencing, both application designers and service providers must know how (in what level) they should provide audio and video services having in mind all requirements for energy consumption and QoE. However, there exists no complete methodology to determine which level satisfies all the constraints regarding energy conservation of handheld devices and the quality that users require. The quality that can be achieved increases with bandwidth and processing power, but so does the cost.

In the rest of this chapter description of the factors involved in the trade-off is provided. Section 2.1 discusses Media compression algorithms examined in this thesis. In section 2.3 QoE and two factors which lead to its degradation will be introduced. It is necessary to mention that term QoE here is limited to Peak Signal to Noise Ratio (PSNR) and Structural SIMilarity (SSIM) as objective measurement methods in video codecs and Mean Opinion Score (MOS) factor as subjective measurement in audio codecs. Finally, discussions regarding energy requirements for transmission and coding phases are provided in section 2.3.

### 2.1 Media Compression

In media transmission over networks, Compression algorithms are able to substantially reduce the amount of information to be transmitted and have been widely used in all kind of media types . Since multimedia systems are highly sensitive to delay and dependent on bandwidth, network condition, computational complexity, and energy resources are highly demanded on these systems. Although compression algorithms reduce bandwidth usage and the amount information to be transmitted, they add additional complexity to the preservation. However, in information which are inherently large (like video), compression is the only viable option for transmission.

Each codec which is used for compressing video information uses an algorithm that removes redundant information from video. This is where different codecs can be compared according to the data rates, output quality and time needed for encoding. These values are dependent to different factors such as the Quantization parameter used in encoding which is mostly shown by  $Q$ . For example, in H.264, which is currently most popular codec in the market, changing the  $Q$  value from 30



to 38 increases PSNR (which is usually expressed in terms of the logarithmic decibel scale) from 30.16 to 35.38. This may seem tenable to have Q equal to 38 which leads to a better output quality and as a result we will have a higher QoE. However, the encoding time also will change from 9.9 seconds to 10.6 seconds (in a 300 frames foreman sequence) which may not be tolerable for users in some applications.

The important point in this regard is that all profiles of H.264 cannot be used in real-time applications in which uploading is being done by low-power, low-computational capable devices. For example, surveillance cameras that are used to upstream live video require a low-complex encoder while the decoder side-which can be a powerful desktop computer might be capable of running more complex decoding algorithms. In these scenarios, H.264 cannot be used. However, in this study, those applications are not investigated and the main focus in video codec is on H.264.

Audio codecs are also used in order to optimize amount of memory used for storing audio data and also avoiding wastage of bandwidth. According to [8] "Codecs are used to convert an analog voice signal to digitally encoded version. Codecs vary in the sound quality, the bandwidth required, the computational requirements, etc." From the G.711 which was introduced in 1972 to G.722 annex B introduced in 2010, different codecs produce different output quality in cost of various amounts of processing power. Each codec must be used in the most appropriate application according to the requirements of the application and also the available bandwidth.

## 2.2 Quality of User Experience

Compression in media application, as discussed in the previous section is viable due to information size reduction, however, it will affect the quality of media negatively considering lossy compression techniques. In this context, it is necessary to mention that this negative influence on quality and as a result on QoE can be divided into two general groups. At one hand, compressing media information degrades QoE directly as the output is not an identical copy of the original information. In this regard, more compression could mean more loss of information and lower-quality output. On the other hand, compressing more data into one single packet which will increase compression efficiency, rises the sensitivity of quality on packet losses. In these scenarios, one single packet loss results in losing more information which might not be tolerable in some applications. That is why we need to know details of compression and why we consider them.

### 2.2.1 Speech compression

Quality of VoIP service is usually evaluated at the application layer using specific tests developed for assessing the perceived speech quality. To perform these tests a

number of methods have been suggested in the past. Basically, we distinguish between subjective and objective tests. Those tests that involve surveying humans are called subjective tests. The metric they operate with is called MOS that provides numerical indication of the quality after compression and/or network transmission. The value of MOS is a number ranging from 1 to 5 with 5 corresponding the best possible speech quality. Objective tests are based on deriving applications layer performance metrics based on codec- and/or network-related performance parameters.

These tests try to provide the required relationship (having good correlation) between system's performance and subjective QoE metric. In order to perform VoIP quality assessment using E-model, it is only necessary to monitor and control a VOIP session to figure out the Packet Loss Ratio (PLR) and then use it to calculate so called "R" value. The latter would be used along with other parameters of speech codecs and will be mapped to a certain MOS score. This score then will be considered as the quality factor. This method will be introduced in more details in chapter three.

Another algorithm which can be used in evaluating the quality of VOIP is called Clarke's model conceived by A.D. Clark [31]. This model can be considered as an improved version of E-model which takes into account the influence of Time Varying Impairments (TVI) and Recency Effect (RE). Packet losses and their distribution over time are not constant and vary during a single session. Distribution of packet losses in a specific session are involved in the Clark's model. Re factor is another important metric whose effect is considered in the overall quality score of a VOIP service.

Recency effect is related to psychological characteristics of individuals which play an important role in subjective quality assessment tests. RE represents experience of a listener of a call quality. In fact, it reflects how a certain user remembers the call quality. Some tests have been done to show the effect of timing of an impairment during a call session on the user's given score to the service. In other words, it is shown that if the call becomes noisy in the beginning of a session, after the end of the session user will give a higher score to the service than the situation in which the same duration of noise happens sometime close to the end of the session. According to [10] a 15 second burst of noise at the beginning of a call resulted in a MOS score of 3.82 while the same burst of noise at the end of call made the MOS score equal to 3.18.

The QoE metric for audio codecs investigated here is this subjective method which are already introduced in the codecs specifications by ITU-T. Relation between MOS grades and quality is as table 1.

Table 2.1: Mean Opinion Score

1	2	3	4	5
Bad	Poor	Fair	Good	Excellent

### 2.2.2 Video Compression

Various compression algorithms for video information have been introduced in the literature among which some are used in realtime applications. They can be either lossy or lossless compression with different number of references. Like any other type of media, compression affect quality of video information in two ways resulting in degradation of quality. These techniques will be introduced and discussed in chapter four in details. Here, the used method for calculating QoE after compression and transmission is discussed.

In order to be able to assess the quality of video perceived by human it is required to run a subjective quality evaluation by asking user's perception of the video they have watched. This method is defined as subjective Video Quality Assessment (VQA). However, in most cases it is not practical since it is highly time consuming and expensive. More over, because of involvement of human and different expectations and ideas of users about video streams it might suffer from lack of accuracy. However, these methods make a valuable database to generate a precise objective evaluation algorithm.

Performing above-mentioned subjective tests consists of a set of steps which must take place. One important task is to choose an appropriate testing system and environment and keeping that condition in a constant level during the whole process of experiment. The reason why this is important relates back to the inherent of video quality assessment in which different parameters are important. For example lightening system of the environment will change the opinion of a user while all other features are kept identical. Source of video must also be chosen and do not change during the whole process of evaluation. This can be done by choosing a number of video sequences and performing tests on them. The order of playing video sequences is also important during assessment and must remain unchanged. Because of wide range of scores given by different users, number of evaluators must not be small or the amusement won't be appropriate.

Considering mentioned issues in subjective video quality assessment, objective measurements are used in this thesis. For objective measurement purpose, two metrics have been considered in this study and each metric has been calculated for four different YUV sequences to ensure integrity of result. These two metrics and their calculation methods are introduced below.

### PSNR

PSNR is mostly represented in logarithmic decibel scale and is used to calculate the ratio between the power of the original sequence and the erroneous sequence after lossy compression process. Higher PSNR value means that coded sequence has a better quality. In order to measure PSNR, one must calculate Mean Square Error (MSE) between original and compressed video sequence first. MSE is defined as:

$$MSE(I, J) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - J(i, j)]^2 \quad (2.1)$$

Then, having MSE for a number of frames of a sequence, PSNR will be:

$$PSNR(I, J) = 10 * \log_{10}\left(\frac{255^2}{MSE}\right) \quad (2.2)$$

$$PSNR = Mean(PSNR(I, J)) \quad (2.3)$$

### SSIM

SSIM is well-known measuring method used to calculate similarity between two images. In objective quality measurement methods, SSIM is the full reference metric which is an improvement to traditional methods like PSNR and mostly considered as a metric which closely represents Human Visual System (HVS) quality assessment [11].

To calculate this metric, according to what has been done in [12] we consider two non-negative images aligned together and name them  $x$ , and  $y$ . let consider  $\mu_x$  and  $\mu_y$  as the mean of  $x$  and  $y$  respectively,  $\sigma_x$  as the variance of  $x$ , and  $\sigma_{xy}^2$  as the covariance of  $x$  and  $y$ . Then three components luminance, contrast, and structure comparison would be given as follows respectively:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \quad (2.4)$$

$$c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \quad (2.5)$$

$$s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \quad (2.6)$$

And finally, SSIM would be:

$$SSIM(x, y) = l(x, y)c(x, y)s(x, y) \quad (2.7)$$

It must be noted that this formula corresponds to single-scale structural similarity

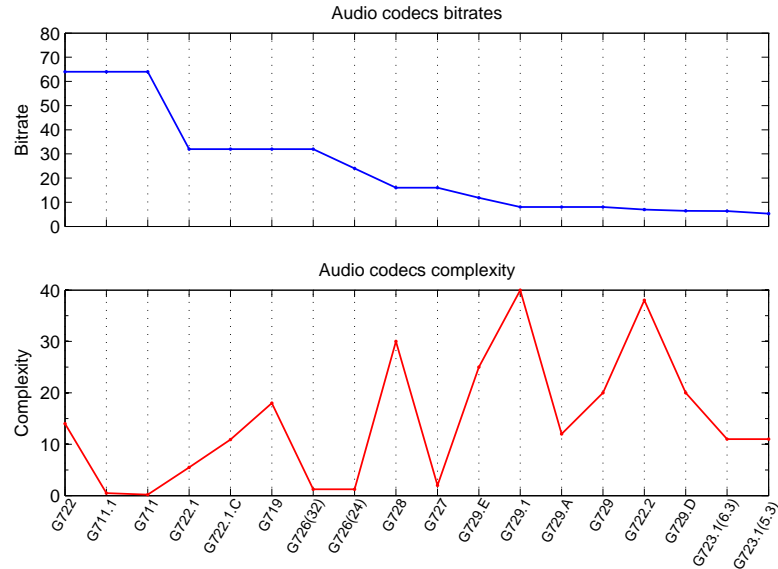


Figure 2.1: Complexity and Bitrate for different audio codecs

and in [13] method for calculating multi-scale SSIM is studied. In this project, PSNR and SSIM for four different sequences foreman, news, Akiyo, and mother and daughter based on 4 different Quantization parameters 20, 28, 30, and 33 are calculated. For this reason, JM v.15.0. which is the H.264 reference codec has been used. More details would be presented in chapter 4.

### 2.2.3 Encoding QoE Degradation

QoE of different audio and video streams would be degraded due to various reasons. The encoding procedure is the first factor affecting QoE of media applications. On one hand, the higher the compression ratio (raw rate to compressed rate) the worse the QoE after compression, e.g. picture or heard quality. On the other hand, this increment in compression ratio directly affects transmission by lowering the amount of information to be sent for the same original data. This trend is known to have complex behavior but in most cases can be sufficiently well modeled by exponential function. A very important effect of codecs is that the more the compression ratio the more information is carried in data units. In network environment it means that the loss of the packet generated by a codec with higher compression ratio will have more negative effect on QoE than that from the codec with lower compression ratio.

This issue is studied for both video and audio codecs. In Figure 2.1 the ratio can be shown by the final output rate of codecs. Data rate here is kbps and complexity of codecs are measured in Million Instructions Per Second (MIPS). It is clear that higher bit rate demands higher bandwidth but less complexity. Encoding is a lossy compression procedure and keeping more part of original data means less compress-

sion and higher QoE. As it can be seen in this figure, the normal trend is that codecs with higher complexity impose less rate on the network. However, at some points this normal trend changes and a codec with higher complexity has slightly higher data rate than others. This feature is because of more complex error concealment algorithms used in the codec which tries to improve QoE. This improvement, requires redundant data to be introduced.

choosing between codecs relates to the level of quality required in the application and energy resource available and the network condition. In poor network conditions, the codec which is chosen might be different with the one chosen in good conditions. For example, more advanced codecs, compress more data in one packet and three or four consecutive loss of packets may result in losing the session. However, advanced codecs, impose less data rate on the network and from this point of view they might be a better option for low bandwidth conditions.

### 2.2.4 Network QoE Degradation

The second factor contributing to end-to-end QoE degradation is the network. There are two major network characteristics affecting the quality of real-time media communication. These are packet loss and delay processes experienced by individual sources. In some applications, there is a strict deadline for packet arrival to the receiver or, if a certain packet is late, it is lost.

As a result, meeting the delay requirements in the network is essential for proper performance of real-time applications. Although real-time applications are highly sensitive to packet delays, the loss process is known to dominate performance degradation. The reason is that the Internet performance in terms of delay highly improved over the last decade and nowadays a plenty of end-to-end paths introduce less than 100ms delay (including those incurred by buffering).

However, even if delay requirements are satisfied, losses may restrain acceptable operation of the service. This is specially important for wireless channels where even in presence of error concealments and channel equalization techniques some erroneously received bits may propagate through the protocol stacks. This causes loss of protocol data units at IP and higher layers. It is known from empirical studies that the QoE degrades exponentially as packet loss rate increases for all media codecs, video, speech and audio ones. The second-order properties of the packet loss process also have profound negative impact on QoE.

In order to understand the impact of network impairment on the QoE we consider H.264 and its reference encoder JM. For this reason it is good to take a brief look at PSNR of the output sequence after 5 percent of packet loss during transmission. The YUV sequence studied in Figure 2.2 is Akiyo and 100 frames are encoded. Out of these frames, 5 percent are dropped and then the output of the encoder is decoded

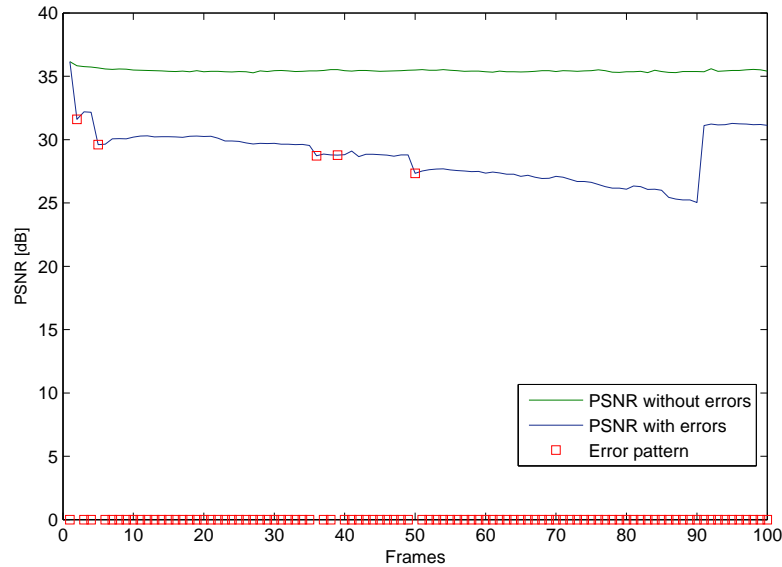


Figure 2.2: Frame-by-frame PSNR for News sequence with  $Q=30$  and 5 percent packet loss rate

again. Using one reference frame its PSNR is calculated. The quantization parameter here is set to 30 and H.264 is in its baseline profile. The green line corresponds to the PSNR of decoded sequence when there is no packet loss. The blue one represents PSNR after packet losses. The order of dropped packets is also important. Consecutive drops will effect PSNR severely while drops with long enough time durations might be concealed in the decoder. The reason is that decoding techniques mostly predict a frame from previously correctly received pictures or future ones. In case of a burst packet loss in highly compressed sequences they won't be able to detect and predict frames. Error Concealment (EC) algorithms are used in codecs to extend capabilities of them in case of these kind of impairments. In JM motion copy and frame copy are two methods used for this purpose. The latter simply copies previous decoded frame to the current lost one but are not able to deal with burst packet losses.

It must be mentioned that EC algorithm used here is set to frame copy which is a built-in option in JM. Although there are more complex EC algorithms that will result in better PSNR in the existance of packet losses, they will introduce more complexity levels. As a result, they consume more energy and are not used in this study. The reason is that, this project mainly corresponds to handheld devices whose lifetime depends on the energy resource i.e. their batteries. In addition to this limitation, their processing power is not enough to process complex algorithms. That is why more complex techniques are rarely used in practice.

## 2.3 Energy Requirements

The amount of energy spent by the client of an application can be divided into the following three components: (i) energy required to maintain the application state, (ii) energy required for encoding and (iii) energy required for transmission. The first component can be considered constant for a given software/hardware configuration irrespective of the type of the codec used for transmission. However, second and third components have variable power consumption.

The reason why energy is important in this context relates to the rapid growth in portable devices. They are capable of connecting to the Internet through cellular networks or other wireless access methods like WiFi. This trend towards multimedia applications using portable devices has been the main focus of many studies in recent years.

Bin et al [14] investigated different requirements for multimedia applications and then categorised them and discussed limitations. These limitations can be the screen size of the handheld devices such as smart phones that is partly studied by [15] and [16]. Processing power limitations, communication issues, and energy requirements are among other source of limitations. In the following subsections, energy related issues are explained in more details.

### 2.3.1 Energy for encoding

The energy spent for encoding varies with the type of the codec and its special features. In audio codecs, depending on the Digital Signal Processor (DSP) used, one can calculate consumed energy for encoding. Investigations done in this project show that the range of encoding energy for these codecs varies from below 1 mW per packet to even 12 mW in some cases. For more details see Chapter 3.

Figure 2.3 shows encoding power of G729 Annex E audio codec released in 1998. This codec uses Conjugate Structure Algebraic Code-Excited Linear Prediction (CS-ACELP) coding algorithm and is presented in red. G722 codec released in 1988 which is a sub-band codec is represented in blue in the same figure. This one divides 16kHz bands into 2 sub-bands, each coded using Adaptive Differential Pulse Code Modulation (ADPCM). It can be seen that energy consumption for all processors in G.729.E is almost double G.722.

In case of video sequences, considering H.264 as an example, the complexity of encoding and, as a result, the energy required for it depends on the targeted resolution and QoE. In this thesis QoE of video sequences will be defined in terms of PSNR and SSIM. The mentioned complexity can be manipulated via a set of internal parameters, e.g. quantization matrix, EC algorithm used, The profile which is chosen for encoding, etc.



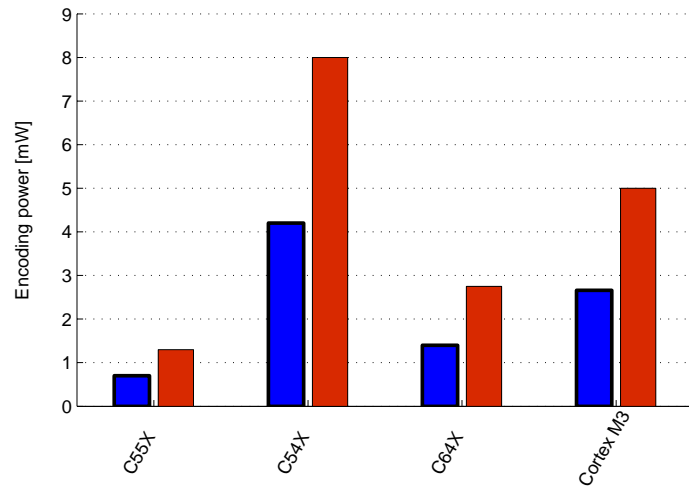


Figure 2.3: Encoding power in mW of G.729.E (red bar) and G722 (blue bar) audio codecs using four different processors.

In this context it must be noted that the higher the resolution and targeted QoE the more energy is required for encoding. However, at the same time the throughput required by the bit-stream increases. The latter dependency is non-linear implying that with the increase of the compression ratio QoE degrades slower compared to the increase in throughput requirements.

### 2.3.2 Energy for transmission

The energy required to transmit a media bit-stream over a wireless channel depends only on the rate of the codec and type of access technology. To clarify, it must be mentioned that Transmission must be divided into two states. That is, each device which support wireless access technologies, has two separate states during its connection to the medium. First state corresponds to the period during which device is sending or receiving data and the other one relates to the time during which the device is sensing and is in its idle state. Most part of the energy consumed for transmission is during active state and device is in very low energy consumption mode while is in inactive state. For different states of transmission energy we have:

$$P_{Tx} > P_{Rx} > P_{idle} > P_{sleep} > P_{off} \quad (2.8)$$

That is why one can assume the energy consumption of transmission is mostly during active state of the device. However, the rate of the bit-stream is a function of the targeted QoE. In principle, the least amount of energy is achieved when the highest possible compression ratio is used and vice versa. Transmission energy can

easily be formulated according to [20] and [21] in terms of  $R$  and  $R_{Max}$  as we can see below:

$$P_{transmission} = P_{active} \cdot \frac{R}{R_{Max}} + P_{inactive} \cdot \left(1 - \frac{R}{R_{Max}}\right), \quad (2.9)$$

where  $R$  is data rate and  $R_{Max}$  is the throughput of the device.

Lower  $R$  means lower transmission ratio and obtaining lower  $R$  is dependent to the codec and its configuration. For example, in JM configuration file one might find some parameters affecting  $R$  using them find the most suitable rate for the application considering other side-effects of that parameter. One might reduce rate but simultaneously double complexity and as a result energy consumption for encoding. In this scenario, instead of any decrement in energy consumption, the final result might show increment in consumption by decreasing rate.

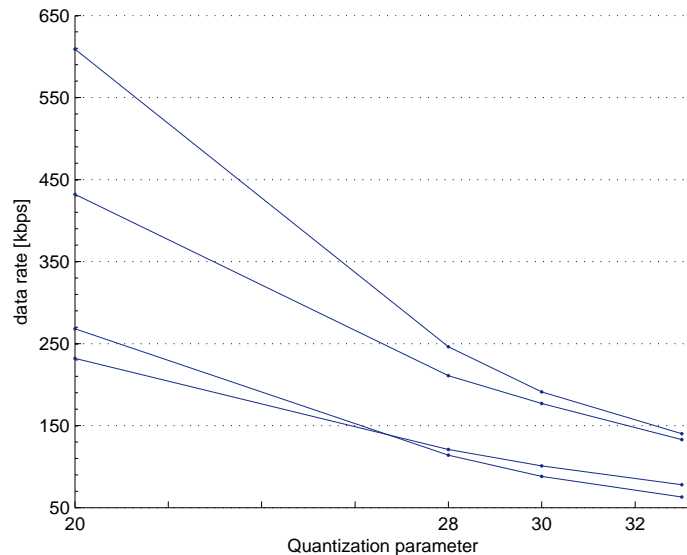


Figure 2.4: Data rate versus Quantization parameter

Another factor which can be used is quantization parameter discussed earlier. Increasing QP value surely decrease output data rate and this decrement in the rate can be considerable as it is depicted in Figure 2.4. There, the effect of QP on data rate is represented for four different sequences all in Quarter Common Intermediate Format (QCIF) whose resolution is 176\*144 pixels. The highest curve corresponds to Foreman sequence and the following curves are for News, Mother-and-Daughter, and Akiyo respectively.

As it can be seen there, increasing quantization steps can decrease the rate and as a result energy required for transmission. However, as it is discussed before, higher QP values result in lower PSNR and SSIM of the output sequence. Having in mind the requirements of the application, one can find the best case accordingly.

### 2.3.3 Energy for different technologies

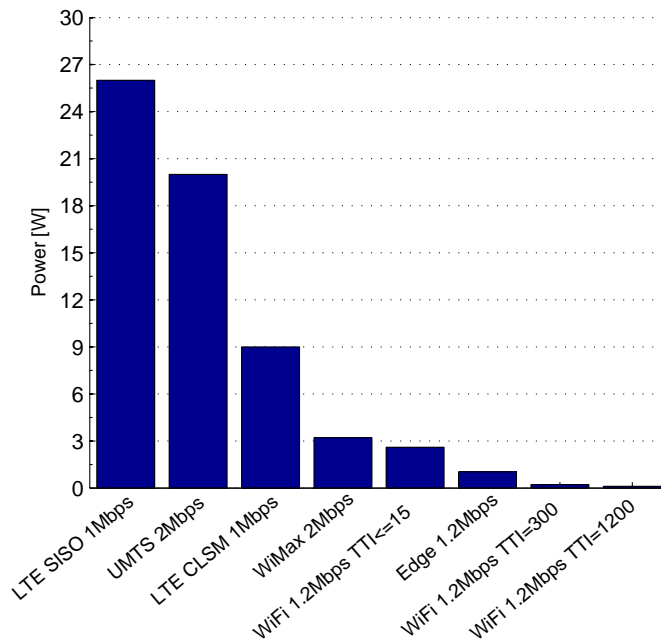


Figure 2.5: Power consumption of different transmission technologies.

Another factor which plays a significant role in energy consumption related studies is the wireless technology chosen for transmission. Although in some situations it might not be possible to choose the technology, the considerable difference in the amount of energy required for transmission per packet, makes it crucial to take this factor into account as well.

The range of transmission energy differs from 26 Watt per packet in 1 Megabits per Second (Mbps) Single Input Single Output (SISO) Long-Term Evolution (LTE) to 0.1 Watt per packet for 1.2 Mbps WiFi while Transmission Time Interval (TTI) is set to 1200s. This huge difference can make a network the main source of energy consumption or an energy efficient one, while producing the same QoE in the same rate. Nowadays, when handheld devices are often equipped with various air interfaces (e.g. WLAN and LTE) choosing the most appropriate technology according to the data rate required and the range of wireless network is one possibility for optimized performance of media services on the move.

Figure 2.5 proves that choosing an appropriate transmission technology, if pos-

sible, would have a non-negligible impact on the total energy consumption. The highest energy corresponds to LTE SISO whose rate is 1Mbps and then Universal Mobile Telecommunications System (UMTS) cellular system whose power consumption is 20 Watt at the rate 2Mbps.

## 3. AUDIO CODECS

The main focus of this chapter is on trade-offs between previously introduced components in audio codecs. The reason why it is important to compare these factors in these codecs and the applications which use them are examined in detail. Since energy required for encoding in case of audio seems negligible compared to energy for transmission, one might say that it is not tenable trading-off them. However, as it will be shown, there are some codecs whose encoding energy is more than energy required for transmission.

This chapter is structured as follow. Section 3.1 introduces VoIP applications and challenges in this area. Different algorithms are used for compression in different codecs. each of them have their own advantageous and disadvantageous. Section 3.2 introduces these methods and compare them. After two primary sections of this chapter, section 3.3 compares all these factors and performs desired trade-off in audio codecs.

### 3.1 VoIP challenges and protocols

This section, tries to study three important factors of VoIP systems with a focus on voice codecs and their features. As it is already discussed, compression, energy consumption and QoE are of high importance. In order to be able to use any VoIP application, one must decide which codec to choose and choosing this codec require precise considerations. Optimizing power usage, bandwidth consumption, and QoE are among these considerations. However, at first it is necessary to see what is VoIP, other alternatives and why to use or do not use VoIP services.

Traditional communication system enabling exchanging voice data i.e. Public Switched Telephone Network (PSTN) and the well-known VoIP are two architectures which play an important role in daily communications. In order to distinguish between advantages and disadvantages of either of these systems, reliability, simplicity of the connectivity, stability, and of course their applications must be taken into account.

Current condition of the Internet and issues regarding mobility, bandwidth available, congestions and losses, and other non-ideal limiting factors make PSTN communications more stable and reliable. On one hand, there is VoIP and its wide range of applications. On the other hand, expensive and time consuming phase of

installing circuit-switched networks and their cabling complexities. all thes, made traditional PSTN companies aware of the need to change [23] or they might suffer from a huge a financial problem.

Various studies is already done to compensate network defects that affect quality of VoIP sessions negatively. Zvezdan et al [24] r investigated PSTN networks and figured out the optimized number of circuits necessary for communications. Using them, they presented a software which is capable of estimating required bandwidth. Yuan [25] introduced an alternative solution to replace H.323 standard to conceal its weak points. There, he presented Session Initiation Protocol (SIP) as a signaling system using a comprehensive study from the features of this protocol to its architecture and inter-working of this VoIP network with traditional PSTN.

### 3.1.1 QoS weak-points

The original application of IP networks was to provide an environment for non-real-time data services. The reason is that these networks operate based on best-effort algorithm. That is why guaranteeing specific quality of service for real-time applications demand additional protocols and considerations. According to this inherent characteristic of IP networks, VoIP providers must introduce and take advantage of new mechanisms. Influencing factors on degradation of quality are listed below:

#### Delay

Delay is one important factor which degrades the quality of VoIP applications and is introduced by both the codec used in the application and impairments of the channel. According to these different resources of delay, one can divide it into network delay, jitter delay, and codecs' processing delay. The latter one, is studied in this project in more details and the result of these studies, is gathered into table 3.1. This table, is a good reference to have the delay of a wide range of ITU-T voice codecs all in one place.

Table 3.1: Processing delay of voice codecs [ms].

G.711.D 12.81	G.722 4	G.722.B 12.31	G.711.1 11.87	G711 0.125	G.718.B 49.62	G.729.1.E 55.69
G.722.1 40	G.722.1.C 40	G.719 40	G.726 0.125	G.723 30	G.728 0.625	G.729.E 15
G.729.1 48.94	G.729.A 15	G.718 42.87	G.729 15	G722.2 25	G.729.D 13	G.723.1 37.5

### **Packet loss**

One problem that considerably affects the quality of service of VoIP applications is packet loss. Although there are some procedures to conceal the effect of packet loss on the quality of the voice, packet loss rates over 10 percent, will affect the quality noticeably. There might be some conditions that they might result in the dropping of the voice session.

Although packet losses may occur due to many reasons, mechanisms used in the Internet these days, makes the probability of these losses happening as a result of noise or other impairments very low. The main source of degradation in QoS due to packet loss is those created by congestion. In order to be able to study the impact of this kind of loss, one must consider the correlation between packet loss and delay on the Internet. The reason of this phenomena is that one sign of possible congestion in the network is increment in the delay (one-way delay) as this shows that there is at least one bottleneck in the network whose queue is getting overloaded.

Some applications have an intelligent system which adapts the rate of the application based on the packet loss rate of the network. Although, this feature might help improve the quality of service considerably, the application must be able to distinguish between different kinds of packet losses. If it is just a random packet loss, decreasing the output rate might not be a good idea. However, packet losses due to congestion are mostly continuous to some extent. In that condition, decreasing the rate is a good way to improve the quality, but in some cases, it might be a crucial approach to keep the session alive.

### **Bandwidth**

Available bandwidth during a VoIP session may vary due to many reasons. Because The consequence of these changes in the bandwidth is that reliability of the quality of VoIP conversations is not guaranteed and change in time. Another effect of bandwidth problems correspond to the protocol used in VoIP applications. UDP does not guarantee delivery of packets and any drop in packet which happen as a result of the available bandwidth affect quality of the voice negatively.

One approach to deal with this problem is to some how reserve bandwidth for VoIP traffic. One of these approaches which has newly been introduced is presented by Luca [27] in which he proposed a bandwidth reservation scheme for VoIP traffic by adopting neural technique. To achieve this reservation, the only item that should be monitored and calculated is packets size and then using some mathematical formulas and implementing the model on some cost-effective devices like FPGAs it is possible to reserve bandwidth. This way, one can improve QoS of VoIP application drastically.

## Security

Security issues are always of high importance for any application in the Internet. In VoIP applications, intruders may be able to listen to other's conversations. On the other hand, they might get access to their voice mail or even find out their ingoing and outgoing calling numbers [28]. For example, one might place a packet grabber on the network while has gained access to the transmission protocol over IP-network and perform eavesdropping as it is done in PSTN systems.

There are a number of security concerns for VoIP applications each of them targeting one or more aspects of the system. Denial of Service (DoS) attacks target Availability of VoIP applications. Getting access to the application and its outgoing and incoming calls might make intruders capable of threatening the confidentiality, integrity and availability of the VoIP system all the same time by making the connection to a third party for example.

In the literature, a number of ways have been investigated to face security issue in VoIP applications. David et al [29] after introducing different attacks in detail and the protocols on which these attacks may impact, presented some solutions for defending security issues. They have shown that one way to make VoIP systems safer is to separate VoIP and data traffic. The other way might be authenticating configuration and/or the signaling protocol and finally, encrypting media might also help improve security of VoIP systems.

## Other factors

In addition to above-mentioned factors which affect QoS of VoIP systems negatively, there are a number of other issues which will have a bad impact on these systems. For example, In poor network conditions, the packets might arrive out of order and this impairment in the delivery of packets might influence the quality of the system. However, this issue might be solved easily as voice packets mostly have a time stamp using that and by using a short buffer list it is easy to reorder packets again.

QoS degradation due to lossy compression algorithms is one thing which must be considered. All voice codecs transform analog signals to digital ones and also decrease their size in order to be able to be transmitted over network. These compression algorithms lower the quality of the output which can be presented by constant  $I_e$  in the E model which will be presented for different codecs in the following sections of this chapter.

### 3.1.2 Standards and protocols

One of the main organizations which standardizes protocols for VoIP applications are Internet Engineering Task Force (IETF) that is an organized activity of the



Internet Society. This community declares how Internet protocols and how the Internet works. The other organization is the ITU which manages global telecom networks and services.

Each VoIP system consists of a number of components co-operating with each other. These components include various gateways to control media and signaling procedures and also a call agent. Synchronization of all these components and handling their issues require different set of rules and protocols.

There are different protocols co-operating to make VoIP services possible. Megaco, SCCP, RVP, MIMO, MGCP are some of these algorithms that are beyond the scope of this project and introducing them might be confusing. However, there are two main protocols i.e. SIP and Session Description Protocol (SDP) which are introduced in this section.

## **SIP**

SIP is a signaling protocol operating at application layer and is client-server based protocol. Using SIP makes end user systems capable of providing a set of services. It might seem so simple for users to start a session with other users by pressing only one button. However, the first problem for session initiation which is solved by this protocol is to map the user's account id which in most cases is user's Email address or phone number with its machine.

SIP does this by sending invitation messages from first user's machine to its provider server and then another message to the second user's server and finally the last invite request to the second user's machine. These messages can be sent either over UDP or Transmission Control Protocol (TCP) and at least must contain the address of the second user and the protocol that makes target machine aware of the parameters of media session which sender requests for the VoIP session. The latter will be introduced in more detail in the following subsection.

Using SIP it is possible to send starting conference sessions, user's multiple devices accessibility for one single user for example from laptop and smart phone, authenticating caller and identifying calling number. Additionally, it enables Internet telephony gateways which are connected to PSTN parties call each other. Another important feature of SIP is its capability to work with other signaling protocols which provide a new set of services.

In addition INVITE messages, there are also a number of other types of messages which makes all these services possible. ACK message is transmitted to inform the other party of the confirmation of last response. To terminate a call it is required to send BYE. If the user wants to cancel ringing before starting conversation, SIP sends CANCEL message. There might be some cases in which having a specific codec or service is not supported by the other party and it is shown in OPTIONS

message used by this protocol. It must be mentioned that all these are in the request messages group. Request messages' structure is shown in table 3.2

Table 3.2: SIP REQUEST message structure.

Method	Request URL	SIP version
--------	-------------	-------------

## SDP

Communication end points need to be at their most effective condition during a communication session. SDP does this by introducing a set of rules in the set up phase of sessions. For this purpose, SDP describes multimedia sessions containing information about session time, data and name, address of end points, and the format of the data. SDP messages are transmitted over UDP and their payload is text.

### 3.2 Codecs compression technologies and their applications

Three factors introduced in Chapter two as the trade-offs would be studied in this subsection for different audio codecs, processors, and transmission technologies. For this reason, ITU-T codecs have been chosen and studied.

Different codecs are intended for different applications based on the coding algorithm used in them and also other factors such as their quality, bandwidth usage, payload size and so on. These main applications of codecs have been gathered and will be introduced in this study. However, they might be used in other applications in some case but those which are provided here are their main usage.

In order to have a better classification, audio codecs are divided into two general groups based on their Packet Per Second (PPS) feature. PPS as it is clear from its name is the rate of generation of audio packets per second and can be calculate as equation 3.1. Using PPS, one can calculate packet duration defined as the time it takes between generation of first bit of packet  $i-1$  and the first bit of packet  $i$ .

$$P_s = \frac{R}{L_s} \quad (3.1)$$

$$P_d = \frac{1}{P_s} \quad (3.2)$$

Where  $P_s$  is PPS,  $R$  is the bit rate,  $L_s$  is payload size, and  $P_d$  is packet duration. Table 3.3 contains codecs whose PPS is more than 50 and table 3.4 corresponds to codecs with PPS over 50. As a general trend codecs which are introduced after the year 2000 have PPS values over 50 and for older codecs, this value is 50 or below. However, G729.A is an exception in this regard.

To be able to summarize technologies in the most optimized way, they are classified into different numbers and the corresponding number of each codec is represented in the table. For this reason, first introduce these numbers as below:

1. ISDN, video conferencing, Voice over Packet Network (VoPN)
2. 3G Wireless
3. VoIP
4. VoATM, ToIP, IP Phone, Private Networks
5. Teleconferencing, Telepresence system
6. WiFi phones VoWLAN, Wireless GPRS EDGE systems, Personal Communications, Wideband IP telephony, and Audio and Video Conferencing
7. multimedia
8. Speech recording and archiving, Digital circuit multiplexing equipment
9. Audio streaming
10. G.729E is a higher rate version of G.729 and is designed to provide higher quality for background noise conditions, music, and tandems. This mode is well suited for music, and it has greater complexity than the original G.729 codecs
11. G.729 Annex A offers the best complexity/quality ratio in the industry
12. Annex D supports multiple bit rates, 6.4/8kbps. Applications include Voice-Over-IP, Voice-Over-ATM, VAD, video conferencing systems, multimedia, store/forward and satellite/wireless communications.

### MDCT transform description

As it can be seen, in the description part of the table there are some new items which must be defined first. Modified Discrete Cosine Transform (MDCT) is a variant of Fourier transform whose difference with other transforms is in the number of outputs after process. MDCT operates on the input and the number of components of the result would be half the input. In addition it must be noted that input must be real number. If we consider  $N$  real numbers as the input, output after transformation will be  $N/2$  real numbers. The formula of MDCT is:

$$X_F = \sum_{n=0}^{N-1} x_n \cos \left[ \frac{\pi}{N} \left( n + \frac{N+1}{2} \right) \left( F + \frac{1}{2} \right) \right] \quad (3.3)$$

Table 3.3: Audio codecs applications and their description with PPS over 50.

Codecs	PPS	Applications	Description
G.711.1	100	3,6	MDCT, A-law, $\mu$ -law
G.711.1.D	240	5	G711.1, BWE, AVQ
G.722.B	800	5	G722, MDCT, BWE, AVQ
G.718.B	56	5	G.718, Sinusoidal coding, MDCT
G.729.1.E	56	5	G.729.1, Sinusoidal coding, MDCT
G.722.1	100	1	low frame loss
G.722.1.C	100	5	MLT
G.719	100	5	MDCT,FLVQ
G.729.A	100	11	CS-ACELP
G.722.2	80	1,2	multi-rate wideband ACELP
G.729.D	80	12	CS-ACELP

### A-law and $\mu$ -law

Analog signals first need to be digitized to be applicable in digital communication systems. For this purpose and to improve the efficiency of linear encoding, a set of methods are available from which we consider A-law and  $\mu$ -law here. In fact these algorithms are used to optimize the digitization process. A-law does this by

$$F(x) = \text{sgn}(x) \begin{cases} K|x| & |x| < \frac{1}{K} \\ \frac{1+\ln(K|x|)}{1+\ln(K)} & \frac{1}{K} \leq |x| \leq 1 \end{cases} \quad (3.4)$$

Where  $K$  is the compression parameter which is equal to 87.7 or 87.6. In addition to this A-law which is used in European countries, there is a  $\mu$ -law utilized in North America and Japan. The purpose of latter is the same as A-law.  $\mu$  law has two types. for the continuous form the formula is:

$$F(x) = \text{sgn}(x) \frac{\ln(1 + \mu|x|)}{\ln(1 + \mu)}, \quad -1 \leq x \leq 1 \quad (3.5)$$

### BWE

Human ear is not equally sensitive to high and low band distortions in the signal. In fact, ear cannot detect distortions in high band very accurately. Advanced codecs, make use of this property to compress digital bits while keeping the quality of encoded audio in a good condition. G.729.1, G711.1 Annex D, G.722 Annex B and some other codecs take advantage of this procedure.

Codecs which use Bandwidth Extension (BWE) encode the lower frequency band of the signal and then using correlations between lower band and the higher one try to predict the latter. Although this might not be a perfect guess, lack of precise detection of higher frequencies in human ears, make this prediction a reasonable

approach.

### **AVQ and FLVQ**

In lossy data compression algorithms, in order to change a multidimensional vector space, which require a high storage space, to a vector with lower dimension demanding less memory to store data, vector quantization techniques are used. For this reason, mostly a codebook is required for this transformation to be done. Algebraic Vector Quantization (AVQ) is a kind of procedure in which the codebook does not need to be stored and as a result the efficiency of this algorithm is even better. Fuzzy Learning Vector Quantization (FLVQ) is another form of vector quantization algorithms in which codebook is obtained through a parameter called gradient-descent learning.

### **Sinusoidal coding**

It is possible to represent an original signal as a summation of various sinusoids. In sinusoidal coding, instead of the signal, parameters of this sinusoid is transmitted and at the receiver side the original signal can be reconstructed. There are a number of schemes for low-rates sinusoidal audio coding both in Time Differential (TD) and Frequency Differential (FD). Different schemes have different delays and frame lengths.

### **MLT**

Modulated Lapped Transform (MLT) is mostly used to make it possible to implement block transform coding in audio compression process in fact it is used to have the analogue audio signal in its frequency domain or time domain representation. In block transformation a DCT would be applied to the data of length M to decorrelate it and then the M coefficients would be encoded.

### **ACELP**

Voice Age corporation introduced a technology called ACELP which nowadays is widely used in different codecs. In ACELP, two different filters are used which are Time Variant (TV). Using these two, speech signals with i samples can be synthesized and scaled. The transform function of two filters in z transform are presented below.

$$\frac{1}{H_1(z)} = \frac{1}{1 - gz^{-t}} \quad (3.6)$$

$$\frac{1}{H_2(z)} = \frac{1}{1 + \sum_{n=1}^N a_n z^{-n}} \quad (3.7)$$

Where  $g$  is the gain parameter which the scaling process would be done according to that. CS-ACELP used in G729 annex A and D is Conjugate Structure-ACELP and is designed to operate with an appropriately band-limited signal [30].

### ADPCM

It sometimes happens that the minimum bandwidth required for transmission must be decreased further while the Signal to Noise Ratio (SNR) remains at the same level. For this purpose, codec must be able to increase quantization step to do so. However, this procedure may affect output quality negatively. This can be achieved using ADPCM.

Getting familiar with different compression technologies used in codecs, it is time to see technologies and the description of those codecs in which PPS is lower than or equal to 50. Table 3.4 presents these codecs and their features.

Table 3.4: Audio codecs applications and their description with PPS lower or equal to 50.

Codecs	PPS	Applications	Description
G.722	50	1	ADPCM
G.711	50	6	A-law, $\mu$ -law
G.726 r32/r24	50	1,2,7	ADPCM
G.728	33	1,3,7,8	G.718, CELP
G.727	50	6	ADPCM
G.729.E	49	10	CS-ACELP
G.729.1	50	3,4	
G.718	50	3	ACELP, MDC, AVQ
G.729	50	1,3,7	CS-ACELP
G.723.1	33	3	

## 3.3 Trade-offs in audio codecs

Previous sections, performed a deep investigation over different aspects of audio codecs.compression algorithms used in a wide range of frequently used codecs and their applications were examined there. This section, studies trade-offs introduced at the beginning of this study i.e. compression, QoE, and energy efficiency.

### 3.3.1 Compression

Different codecs encode and produce voice data at different rates. Higher complexity of codecs might result in lower data rate which reduce energy required for transmis-

sion. However, running those codecs will increase computational complexities which means higher processing energy. On the other hand, compression algorithms used by codecs have a direct impact on the QoE of the output as well.

Comparing codecs compression ratios makes it necessary to consider other factors than data rate as an indicator of compression efficiency as well. Payload is one of those factors. Payload of the codecs is introduced both in milliseconds and bytes. In addition, calculating bandwidth consumption is another process which would be possible when the payload of the voice data is in hand. Band width calculation would be different as different technologies impose their own header. The general equation which is used here for calculation of bandwidth is presented below:

$$S = H + P \quad [Bytes] \quad (3.8)$$

$$B = S * PPS \quad [KBps] \quad (3.9)$$

Where H is header size and P is the payload of voice packets. In formula 3.9, B denotes bandwidth and S stands for voice packet size. the total voice packet size can be calculated knowing that the packet is transmitted over Ethernet normally or, for example, Compressed Real-Time Transport Protocol (cRTP) is used in transmission. It must be noted that cRTP is not an option while using Ethernet. in normal transmission, the IP/UDP/RTP header would be 40 bytes as follow:

1. IP header=20 bytes (20\*8=160 bits)
2. User Datagram Protocol (UDP) header=8 bytes (8\*8=64 bits)
3. Real-Time Transport Protocol (RTP) header=12 bytes (12\*8=96 bits)

However, there are some cases in which cRTP is used. Using this protocol reduces the 40 bytes of overhead to only 2 bytes. As mentioned before, this technology cannot be used over Ethernet. Another protocols which might be used here are Multi-link Point-to-Point Protocol (MLP) or Frame Relay Forum (FRF). They will add 6 more bytes as layer 2 header. In case of Ethernet, this will be 18 bytes of layer 2 header which also includes Frame Check Sequence (FCS) or Cyclic Redundancy Check (CRC) headers.

To conclude the voice packet size used in the formula 3.8 which is the total packet size, can be calculated using the formula below. According to the type of protocol used, layer 2 header and/or IP/UDP/RTP header sizes might change. As a result the amount of bandwidth usage per call for different codecs in kbps can be calculated as:

$$H = \text{Layer2header}(MP \text{ or } FRF.12 \text{ or } ethernet) + (IP/UDP/RTP) \quad (3.10)$$

$$\text{Layer2header}[\text{bytes}] = \begin{cases} 6 & MP \\ 2 & FRF.12 \\ 18 & Ethernet \end{cases} \quad (3.11)$$

$$\text{IP/UDP/RTPHeader}[\text{bytes}] = \begin{cases} 40 & Ethernet \\ 2 & cRTP \end{cases} \quad (3.12)$$

Considering different cases of formulas presented above, bandwidth usage can vary in a wide range. Figure 3.1 depicts the differences in bandwidth usage using different protocols. This study, investigates almost 20 different voice codecs. However, the payload size for these codecs changes between 10 bytes to 80 with steps of 10 and there are also two 160-bytes-payload codecs. This figure takes into account all these values of payloads regardless of the codecs. PPS values are taken from Tables 3.3 and 3.4.

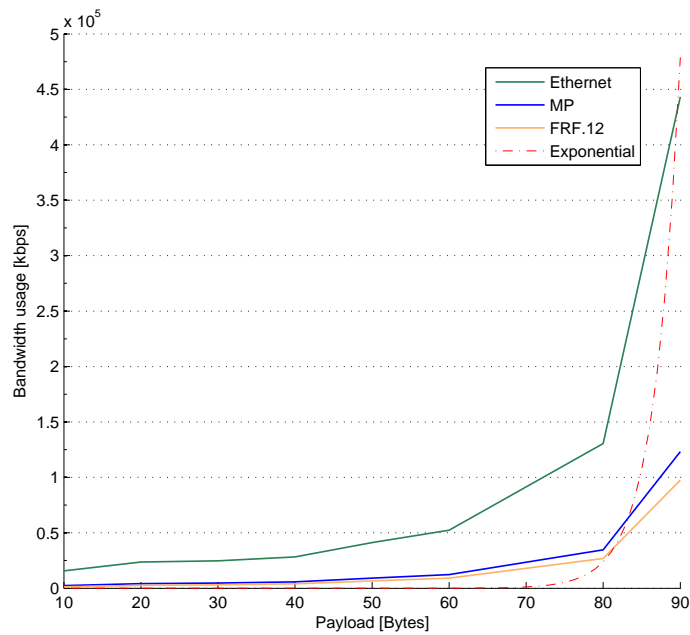


Figure 3.1: Bandwidth usage of audio codecs.

Figure 3.1 shows that using Ethernet, imposes highest bandwidth usage and FRF.12 has the lowest usage as the amount of overhead it adds to voice payload is much lower than Ethernet. All these three curves are some how similar to exponen-



tial curve (red one in the figure) and their increment is almost exponentially based on the payload size of the codec.

Table 3.5: Audio codecs Bitrate, payload, Ethernet bandwidth required.

Codecs	bitrate [kbps]	payload [Bytes]	payload [ms]	Eth. Bw [kbps]
G.711.1.D	96	50	5	245.76
G.722	64	160	20	87.2
G.722.B	64	10	5	563.2
G.711.1	64	80	5	126.4
G.711	64	160	20	95.2
G.718.B	36	80	20	71.1
G.729.1.E	36	80	20	71.1
G.722.1	32	30	20	86.4
G.722.1.C	32	30	20	86.4
G.719	32	40	20	94.4
G.726	32	80	20	55.2
G.726	24	60	20	47.2
G.728	16	60	20	31.5
G.727	16	40	20	47.2
G.729.E	11.8	30	20	43.2
G.729.1	8	20	20	39.2
G.729.A	8	10	20	39.2
G.718	8	20	20	39.2
G.729	8	20	20	31.2
G.722.2	6.60	10	20	38
G.729.D	6.4	10	20	56.32
G.723.1	5.3	20	30	21.9

Table 3.5 presents all values of different codecs payload sizes both in second and bytes. It also the rates of codecs plus bandwidth required per call over Ethernet. The point is that some codecs have variable bit rates which this also improves their capabilities in lowering bandwidth usage. However, this table only contains one of their data rate and the corresponding bandwidth is based on the bitrate written in the table.

### 3.3.2 QoE

MOS defined in ITU-T is a subjective measurement of quality of VoIP. However, subjective inherent of this kind of evaluation makes it expensive and time consuming to be implemented over different codecs. MOS of different audio codecs is presented here as one factor related to QoE of codecs and is taken from codecs' specifications.

In addition to this subjective evaluation method, a number of objective methods are also standardized in recent years. Among them Perceptual Speech Quality

Measure (PSQM), Perceptual Evaluation of Speech Quality (PESQ), Perceptual Analysis Measurement System (PAMS), and the well-known E-model presented in ITU-TG.107 are the most important ones.

PEQM, PESQ, and PAMS are not basically designed for data networking applications. As a result, they are not an accurate approach for assessing call quality in VoIP applications. In fact, they might evaluate the quality of voice well but in order to examine the condition of network during which that quality is achieved is not a built-in option of these methods. One might use external devices and hardware to measure some parameters while using these methods. Parameters such as network delay, jitter delay, packet loss, and so on.

Another important factor while choosing an assessment method is the output of the evaluation. It is highly desirable to have a single output such as a single number that includes all of the influencing parameters. Then, it can be referred as the quality of voice instead of a set of numbers and parameters which might be confusing.

That is why E-model plays an important role in all VoIP evaluation methods. The output of this model is a single number called "R value" which lies between 0 to 100. Using the final number obtained from this model, one can easily give a comment about the quality of voice session. This model contains different parameters in its equation. The basic formula is presented below:

$$R = R_o - I_s - I_d - I_e + A \quad (3.13)$$

Where  $R_o$  examines the effect of signal to noise ratio on the quality and  $I_s$  shows how impairments happening simultaneously with speech signal affect quality.  $I_d$  considers negative effects of delay happened in the network on quality.  $I_e$  corresponds to negative effects of equipment and finally  $A$  is the advantage factor. After subtracting all impairments from signal to noise ration  $R_o$ ,  $A$  value would be added to the result to compensate it. The reason of this factor relates back to the fact that sometimes due to network conditions, user might prefer to keep the voice session even with poor quality instead of dropping it. There might be regions which are hard to reach or in case of mobile devices, this advantage factor can be helpful.

However, here another equation of E-model simplifying it to the effect of used codec on the quality is considered. For this purpose,  $R_o - I_s$  can be assumed to be equal to 94 [31] which will result in:

$$R = 94 - I_d - I_e \quad (3.14)$$

$I_{e,eff}$  which is the effective equipment impairment factor tells how low bit-rate codecs may affect negatively quality. More over, the impact of packet losses whose

Table 3.6: Audio codecs MOS and  $I_e$  value.

Codecs	MOS	$I_e$
G.711	4.3	1
G.729.E	4.2	4
G.711.1	4.17	0
G.722	4.13	16
G.722.1.c R48	4.10	-
G.722.1 R32	4.09	-
G.722.1 R24	3.98	12
G.729	3.92	10
G.723.1 R6.3	3.9	15
G.726 R32	3.85	7
G.727 R24	3.83	25
G.722.1.c R32	3.80	-
G.729.a	3.70	20
G.723.1 R5.3	3.65	19
G.728	3.61	7
G.729.d	3.60	20
G.722.1.c R24	3.52	-
G.726 R24	3.51	25
G.727 R16	2.84	50

distribution is random is also included in this component. However, in order to be able to compare different codecs, it is only necessary to know what is the value of  $I_e$  even without taking into account packet losses. These values for different codecs are given in ITU-T Rec. G.113.

The effect of different codecs on the voice quality according to values provided in G.113 recommendation are given in Table 3.6. It must be mentioned that they don't take into account packet losses. In addition to that, MOS value for these codecs are also provided in the same table.

Considering a threshold for MOS value that makes it possible to communicate, one can compare different codecs in terms of their QoE easier. Figure 3.2 considers 3 areas of quality evaluation. One is Excellent region which is for MOS values over 4. The other one, acceptable/ Good region corresponds to those codecs whose MOS is between 3 and 4. Finally, unacceptable part which contains only G.727 with data rate 16 kbps whose MOS value is below 3.

### 3.3.3 Energy Consumption

To be able to have a good comparison between energy consumption and other factors which are introduced in previous subsections, it is better to divide energy consumption into two different parts and examine each of them separately. One of these two

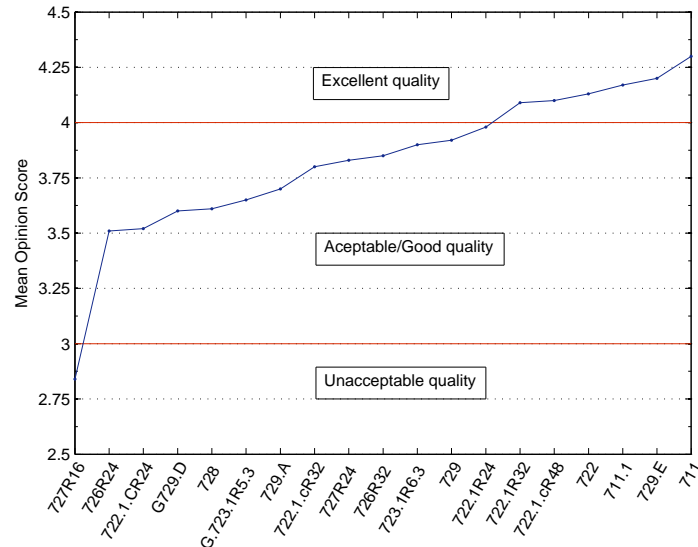


Figure 3.2: Dividing MOS of audio codecs among three different regions of quality.

concepts are energy consumed for encoding which relates to the processor used in DSP. The other one is energy required for transmission. Further in this study, a comprehensive table would be presented taking into account different transmission technologies.

Figure 2.3 proved that energy for encoding using a specific processor may be almost 8 times for the same codec in another DSP. Although choosing between DSPs require considerations more than only energy consumption, they are beyond this study and here the only important thing is power consumption.

On the other hand, in Figure 2.5 it is shown that the communication technology plays a non-negligible role in the energy required for transmission. Using 4G LTE has its own benefits. however, in that figure it can be seen that LTE SISO with 1 Mbps bandwidth would consume almost 25 times WiFi 1.2 Mbps whose TTI value is set to 1200.

Considering above-mentioned facts proves that it is not convenient to examine only QoE or compression efficiency of codecs. One of existing correlations correspond to compression ratio. Good effect of higher ratio is in using bandwidth efficiently. However, as it is already mentioned, in poor network conditions, losing more compressed packets means more loss in information. This is even more important examining handheld devices where lifetime of device depends on the battery. Choosing a wrong codec or an inappropriate technology may result in wastage of battery and make the device working hours without recharging noticeably low.

### Encoding power

Four different DSPs have been examined in this study. At first it was necessary to have an estimation of codecs processing requirements. Most of voice codecs announce their processing requirements in MIPS, Million Operations per Second (MOPS) and Weighted Millions of Operations Per Second (WMOPS).

Although going deep in the definitions of these three might be confusing and certainly beyond this study, an introductory understanding of them is necessary. As already mentioned MIPS stands for Million Instructions per Second which maps each operation to an instruction. However, This is not true when parallel machines are used in the application. In that case each instruction has more than one operation and MOPS must be considered. In the first case (risk machines) both MIPS and MOPS will have a same result.

Knowing the processing measurement unit of different codecs is the first step. Then, it is required to find energy consumption in mW or W and map MIPS or MOPS to them. Most used DSPs' data sheets are the best reference for this purpose. For this reason, four different DSPs have been chosen .According to their data sheets the amount of energy they require per MIPS have been extracted. Used processors are listed below:

1. C55x power consumption rate=0.05 mw per MIPS
2. C54x power consumption rate=0.3 mw per MIPS
3. C64x+/66x/674x/64x power consumption rate=0.1 mw per MIPS
4. Cortex M3 power consumption rate=0.19 mw per MIPS

Table 3.7 represents processing complexity of codecs in MIPS along with their corresponding encoding power in different DSPs. As it can be seen, C55x which is from Texas Instruments C5000 series, has the lowest power consumption in comparison with others[32].It is assumed that ARM 7 processors are used in combination with c500 DSP series. For C6000 series, mostly ARM 9 processors will be used. TMS320C6x is an example of c6000 series DSP and TMS320C55x is an example of C5000 family.

### Transmission power

It is mentioned that trading power consumption is important and one component of energy consumption resources have been investigated above. however, another important factor in this context is energy required for transmission which would be discussed in this part.

Table 3.7: Energy required for encoding using 4 different DSPs.

Codecs	MIPS	C55x	C54x	C64x+/66x/674x/64x	Cortex M3
G.722	14	0.7	4.2	1.4	2.66
G.711.1	0.5	0.02	0.15	0.05	0.09
G.711	0.2	0.01	0.06	0.02	0.03
G.722.1	5.5	0.27	1.65	0.55	1.04
G.722.1.C	10.9	0.54	3.27	1.09	2.07
G.719	18	0.9	5.4	1.8	3.42
G.726	1.25	0.62	0.37	0.12	0.23
G.728	30	1.5	9	3	5.7
G.727	2	0.1	0.6	0.2	0.38
G.729.E	25	1.25	7.5	2.5	4.75
G.729.1	40	2	12	4	7.6
G.729.A	12	0.6	3.6	1.2	2.28
G.729	20	1	6	2	3.8
G.722.2	38	1.9	11.4	3.8	7.22
G.723.1	11	0.55	3.3	1.1	2.09

Estimating energy consumed in transmission depends on many factors from the range of coverage to data rate, transmission power shown by  $T_x$ , node B, and so on. However, this study takes into account an average but convenient estimation for different technologies. Those technologies which are studied in this paper and a brief definition of them are listed below. These are power consumed in the sending phase.

1. LTE 1 Mbps SISO and LTE 1 Mbps clsm. It must be mentioned that LTE is the newest wireless broadband technology whose access technology is SOFDMA which is multiple access which allows LTE to support variable bandwidths.
2. WiMAX which is intended for broadband communication and standardized in IEEE 802.16. Fixed WiMAX uses Orthogonal Frequency Division Multiple Access (OFDMA). For mobile communications it would be 802.16e and uses SOFDMA.
3. UMTS which is developed by European Telecommunications Standardization Institute (ETSI). For mobile voice and data UMTS is a good solution. Its multiple access technique is based on Wide-band Code Division Multiple Access (W-CDMA)
4. EDGE allows people to have a technology which is backward compatible with old GSM system while providing higher transmission rates.
5. WiFi is a synonym for WLAN IEEE 802.11 standard. WiFi can be used both

Table 3.8: Energy required for transmission over different technologies [mw].

Codecs	LTESISO	LTECLSM	WiMAX	UMTS	Edge	WiFi TTI=1200
G.722	49.5	17.13	3.04	19.04	1.63	0.16
G.711.1	32.86	11.37	2.02	12.64	1.08	0.1
G.711	49.50	17.13	3.04	19.04	1.63	0.16
G.722.1	22.46	7.77	1.38	8.64	0.74	0.07
G.722.1.C	22.46	7.77	1.38	8.64	0.74	0.07
G.719	24.54	8.5	1.51	9.44	0.81	0.79
G.726	32.86	11.37	2.02	12.64	1.08	0.1
G.728	28.7	9.93	1.76	11.04	0.95	0.09
G.727	24.54	8.5	1.51	9.44	0.81	0.79
G.729.E	22.46	7.77	1.38	8.64	0.74	0.07
G.729.1	20.38	7.05	1.25	7.84	0.67	0.06
G.729.A	18.30	6.33	1.12	7.04	0.6	0.06
G.729	20.38	7.05	1.25	7.84	0.67	0.06
G.722.2	18.30	6.33	1.12	7.04	0.6	0.06
G.723.1	21.21b	7.34	1.3	8.16	0.7	0.07

in small homes, apartments, and hospitals and in a wider range like cities. For example Sunnyvale, California, provides a city-wide free WiFi

Calculations of energy required for transmission in this study is based on the amount of power that each of these technologies use in a specific data rate. For example LTE uses 26 watts in 1 Mbps throughput and in single input single output mode. On the other hand, the bit rate of each of these codecs is known to us. Using a simple math equation, one can find transmission energy consumption estimation.

Having transmission energy and encoding energy calculated in the previous subsection, one would be able to compare these two values. In some case encoding power might be even more than transmission power while in other conditions or in the same condition using another technology, this might change.

An example to this comparison of power consumption in encoding process and transmission phase one can consider G711 and G722 codecs. In both of them transmission power over WiFi 1.2 Mbps whose TTI is equal to 1200 is 0.16 mW while encoding power for G.711 is 0.01 mw which is lower than transmission and for G.722 it takes 0.7 mw to encode which is more than transmission.

Table 3.8 shows all these transmission powers for different codecs. For codecs whose data rate is the same, in transmission power would also be the same which is shown in the table. It must be mentioned that all these values for transmission power are calculated in mW and have been put in the table.

## 4. VIDEO CODECS

The process of encoding, transmission, decoding, and finally playing back of video sequences is more energy consuming and more complicated than what is discussed in audio codecs. Higher complexity of video codecs imposes more energy consumption in the phases of encoding and decoding than audio codecs. On the other hand, transmission energy in video related applications is mostly more than those applications used for audio transmission. The main focus in this chapter is on these issues and a solution to minimize energy consumption in both phases of coding and transmission and is structured as follow.

Section 4.1 introduces the whole structure of video content communications and compression techniques used in this context and its challenges. The standard used in this thesis is H.264 which is widely used these days. However, this standard has different profiles for different applications and cannot be used interchangeably or might not work properly. These issues are discussed in details in section 4.2 and its subsections. Section 4.3 defines trade-offs for video content communications using H.264 and also provides a method for minimizing energy consumption while keeping the quality of video in an acceptable level.

### 4.1 Video communications

#### 4.1.1 Monitoring quality of communication

In the context of multimedia communications, there still exist some problems in both well-known best-effort Internet and also cellular networks which affect these services negatively. The reason of these technical problems is inherent features of these communication infrastructures and restrict demands of real-time applications. These requirements of video communications result in their dependency on a minimum level of QoS to be able to perform appropriately. For example, their high sensitivity to losses and also timely delivery constraints must be strictly satisfied.

These issues cannot be handled in best-effort communication infrastructures without collaboration of extra protocols and standards. Using a set of protocols, one can monitor network condition and the application demands so dynamic management of network resources could be possible. One of these procedures for traffic handling is taking advantage of RTP/RTCP while transmitting over UDP.



Real-time Transport Protocol (RTP) is designed to standardize packet structures in multimedia services over IP networks. Additionally, RTP Control Protocol (RTCP) is used in conjunction with it to provide monitoring capabilities as well. RTP defines the format of packets and these packets are sent through IP networks and then RTCP is used to monitor network performance and its QoS.

In order to deal with mentioned problems of real-time data transmission over best-effort communication architectures, RTP packets consist of some additional flags and elements. Timestamps are used for synchronization purposes. To provide the possibility of detecting packet losses occurred during transmission, sequence numbers are also included into these packets. Format of the data which is being carried by the packet presents in each packet so at the receiver side, appropriate process can be executed. After a number of transmission and in a periodic form, RTCP exchanges controlling information between sides of communication so that monitoring and handling traffic of network will be possible.

Monitoring process must be done on all entities of a network involved in the specific session so that each one can dynamically handle possible impairments. For this purpose, whenever sender gets back network information from receiver(s) it spreads that to all of other entities. This way, each component has an updated version of those parameters which are being monitored. It must be noted that RTP mostly owns 95% of the entire traffic while RTCP bandwidth usage is limited to 5% of the whole traffic.

### 4.1.2 Compression challenges

Digitization of video content has a wide range of applications each of which require different services and different levels of those services. For example, some of these applications are security purposes monitoring, digital televisions, DVD, video recorders, and streaming video over the Internet. In each of these applications different codecs or different profiles of a same codec might be used.

The reason why codecs and their compression algorithms are used relates back to the storage capacity requirements of uncompressed video information and the required data rate for raw video after digitization. For example a typical NTSC video to be digitized at 30 frames per second requires 110 Giga Bytes to be stored and 165 Mbits per second of data rate.

These storage requirement and data rate makes it almost impossible to use video applications without using an appropriate compression algorithm. Codecs use different compression techniques each of which tries to reduce data rate and storage capacity while keeping the visual quality at its best level possible. It's almost impossible to have the same quality of raw video since most of these compression techniques are lossy.

Although a specific codec might be the best choice for one of these applications, it might be impossible to use that in other application. For example, H.264 High Profile (HiP) which is used in Blue-ray discs, cannot be used for real-time streaming over the Internet. The reason is that, it imposes a high complexity level and takes relatively high processing time which might be intolerable in live video streaming applications.

Choosing a compression scheme for a specific application requires more detailed considerations. One of these considerations is the demand level of the application. If the image quality after compression be lower than the minimum requirement, that algorithm would fail. More over, the display size is another important factor in this context.

In addition to the minimum quality, maximum bit rate of the application is also important. Another fact to be considered is the storage capacity. It might be possible to reach the minimum quality and desired resolution in cost of expanding size of the compressed video. This would require more capacity and as a result bigger buffer.

Maximum latency of the codec which can be tolerated by the application is also important. All these issues must be taken into account before a specific codec is chosen. Different codecs handle these issues using their defined protocols and trade off between encoding/decoding devices cost, desired bit rate and storage size and other important facts.

Although downloading a whole video information like an entire MPEG-4 movie has been a robust way of using the Internet for multimedia applications, streaming is becoming more and more popular these days. Conventional downloading approach demands a high buffer space and user must wait for the information to be downloaded entirely which may take from some minutes to even hours.

Streaming, on the other hand, solved these kind of problems and is being used by many users as it is more flexible than downloading. In this approach, instead of waiting for the whole sequence to be received, each video stream is divided into chunks. These chunks are called packets and then these packets would be encoded at the sender side and transmitted over the network. At the receiver side, these received packets would be decoded and then can be played back.

The process of sending packets at the sender would be on-going simultaneously while the receiver is decoding and streaming already received parts of the video. Although this approach requires less buffer space than downloading, many issues are also introduced. For example the process of encoding, transmitting, decoding, and playing-back that must be done all together would result in a kind of delay problem called low delay. An example of low delay problem can be seen in IPTV applications where user wants to change between channels at any time and this

amount of delay must be minimized.

One can distinguish between multimedia communications in a number of ways. One of these classifications takes into account the number of senders and receivers and its components are described as below:

1. Unicast in which one sender and one receiver are connected like media on demand applications
2. Multicast in which there is one sender and a number of receivers. Two types of multicast communications are IP multicast and P2P multicast which is a application level multicast. In the context of bandwidth utilization, multicasting is much more efficient than multiple unicasts.
3. Broadcast in which sender will find all the receivers to whom it is connected and then transmits to them.

Communication channels do not have a steady state and because of many things their conditions vary during a multimedia session. That is why, having the ability to monitor channel to see its impairments or improvements is always a plus. This can be done only in unicast communications where one sender and one receiver are connected. In addition to monitoring channel, this feedback makes it possible for the receiver to tell the sender about end-device characteristics as well.

All this information helps sender to decide which codec or codec's profile is the best choice accordingly. Sender has the possibility to adapt the compression ratio, data rate generation, error correction algorithm and other features to optimize the quality of the video received and seen by receiver. This will directly affect QoE and power consumption.

Communicating video content multimedia consists of three main parts each of which includes one or more components. These parts are sender side, transmission medium, and the receiver side as shown in Figure 4.1.

## 4.2 H.264

In the context of media data, terms codec and container are confusing for many people. A compressor-decompressor or in its shortened form codec, is responsible for either compressing raw media data (which then can be transmitted or stored somewhere) or decompress them (to be viewed at the receiver side after transmission or to be transcoded). Containers are used to hold various media formats. Since different codecs are used for encoding (compressing) raw media data, containers must be able to work with different coding schemes. The more coding schemes a container can handle, a better one it is.

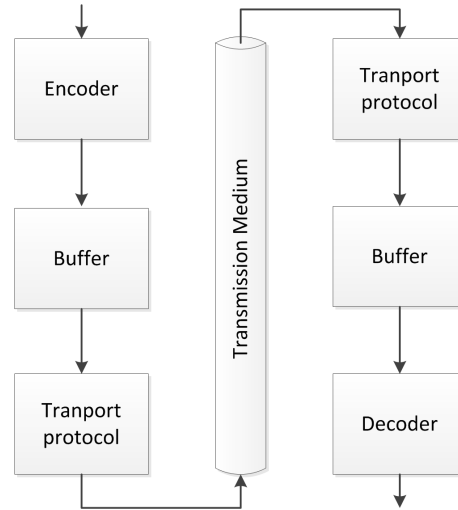


Figure 4.1: Video communication system.

Containers are used in different applications. At this point we introduce a number of most common containers which are wellknown. Advanced Systems Format (ASF) is a Microsoft-based container file which can have extensions like .wmv, .asf, and .wma. The latter, for example, is compressed by Windows Media Audio codec but is an ASF container file. Audio Video Interleave (AVI) is another wellknown codec which is defined by Microsoft. Motion Pictures Expert Group introduced widely used MP4 container. H.264 codec is used for compressing video data and AAC codec for audio data inside MP4 files. Other commonly used containers are Flash, Matroska (.mkv extensions) and DivX.

Two main standardization organizations work in video codecs field which define and provide different standards and protocols. ITU organization which is already introduced in this study and all H.26x standards are defined by ITU. The International Standards Organization (ISO) is another standardization body which has introduced well-known JPEG and MPEG standards. JPEG is used for still images compression and MPEG is used for moving pictures. This thesis, however, considers most recent and powerful codec. H.264 standard is the latest generation of video codecs which is nowadays widely accepted and used in different applications from streaming to recording [17], [18], [19].

### 4.2.1 Codec Structure

H.264 improves compression efficiency by reducing data rate around 2 times versus Mpeg-2 and MPEG-4 simple profile. This huge improvement in compression and data rate, makes many on demand applications available over the Internet a well noticeable reduction in the memory space required to store an HD movie.

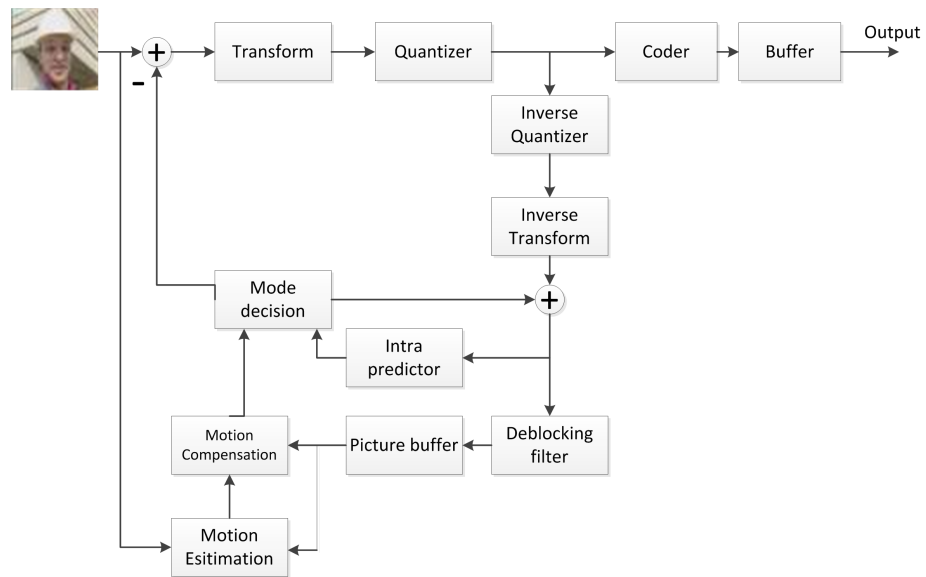


Figure 4.2: H.264 structure, encoder side.

## Encoder

In order to have a better view of this codec, it is necessary to take a look at the structure of the codec. For this reason, Encoder and decoder parts of the codec are separated and the encoder part is depicted in Figure 4.2. Different components of this block diagram are explained below.

1. **Transform** used in H.264 is a  $4 \times 4$  integer transform with specified coefficients make it possible to be invertible. In this coding algorithm, values of different pixels in both cases of Inter Macro Blocks (MB) or intra MBs are predicted. In the case of Intra MB these values would be predicted from neighboring pixels of the same picture while in intra mode, previously decoded reference pictures would be used for this purpose.
2. **Quantizer** which is also called scaling is the next block after transformation. Based on the quantization parameter of the intended MB, scaling factor would be chosen. Each MB has a number of sub-blocks within which the element to be scaled can be found. The position of this element also affects the scaling factor value. It also might be possible to use a feed back from Buffer block to Quantizer to enable rate controlling procedure. In that case, the algorithm which is used for rate controlling purpose chooses the quantization parameter.
3. **Coder** After first step, transformed coefficients would be quantized at the quantizer. After all these, there is a need to a block to code these transformed coefficients which happens at the coder block. Two different coding schemes are used at this block which are Context-Adaptive Variable-Length Coding

(CAVLC) and Context-Adaptive Binary Arithmetic Coding (CABAC) both of which are which is a lossless compression algorithm. Both of these coding methods are supported in main profile of H.264.

4. **Inversing blocks** The output of Quantizer would be divided into two branches. One of them would be passed to the entropy encoder which then will be buffered and transmitted over channel. The second branch, must go through an inverse transformer to derive (after inverse quantization) primary data. That is why being perfectly invertible is important in transformation process.
5. **Mode decision** It is necessary to choose coding must be done in inter or intra mode which is why this block is placed. According to the rate and distortion and also the algorithm used in rate control, the most efficient coding mode would be selected at this stage.
6. **Deblocking filter** One feature of H.264 for is its possibility of using multiple reference pictures. The task of this filter is providing those references for the pictures of a sequence which are intended to be used as reference pictures. This filter operates on both MBs coded in inter mode or intra mode. In case of inter mode coding, it will be applied to the picture after motion compensation part. For Intra mode coding, it will be applied after intra mode prediction.
7. **Other blocks** Intra prediction block is responsible for predicting intra MBs as it is clear from its name. However, this process is a complex one which is beyond this project. Motion estimation and motion compensation blocks, as it can be understood from their names, are used for predicting inter mode coding scheme.

### Encoding overview

In encoding process, encoder must distinguish between header and MB elements. These two parts would be encoded using different coding algorithms in H.264. MB level coding, as it is mentioned previously, the first thing that encoder must decide is the entropy coding type to be in intra or inter mode.

Then, According to the macroblock coding scheme which is decided, encoder must be able to perform rate controlling techniques. The reason why this functionality is important relates back to its significant effect on the output quality. In other words, the target of this part is keeping coding bit-rate in the range of maximum bit-rate while keeping the quality of video in an acceptable condition.

Slicing elements of the input video sequence and coding each element using appropriate algorithm and also performing bit allocation on MBs which at this step are partitioned into different blocks, it's time to reconstruct MBs again. At this step

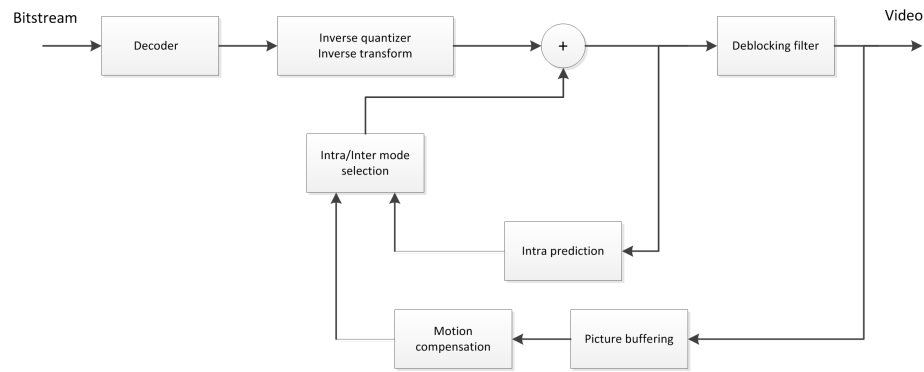


Figure 4.3: H.264 structure, decoder side.

MBs are divided into two groups labeled as coded or not coded. If it's not coded it will directly be passed to prediction block to be reconstructed. If Mb is labeled as coded, it must first go through inversing blocks to become like a not coded one and then be passed to prediction blocks.

## Decoder

The block diagram of decoder can be seen in Figure 4.3. Components of decoder are the same as what we saw in encoder structure. First block is entropy decoder which distinguishes between coded MBs in CAVLC or CABAC and decodes residual signal in a n appropriate way.

After entropy decoding process, inverse blocks operate on the decoded MBs and according to the mode of coding which can be either intra on inter, the residual block can be reconstructed. This will be done for all blocks of a MB and then for all macroblocks of the picture.

### 4.2.2 Profiles and levels

It is already mentioned that H.264 is one of the most widely used standards these days meanwhile it is a very complex standard and using it appropriately requires a good knowledge of its features and the requirements and limitations of the specific application for which this standard is intended to be used. Two main parameters of the codec are profiles and levels each of which address different problems.

## Profiles

Profiles of H.264 can be used to control different complexities and power consumption due to processing. Different profiles use different encoding techniques for compression. In addition to complexity and power, another trade-offs of these profiles are quality enhancement and the complexity of decoder. More complex profiles

take advantage of more advanced compression algorithms which require powerful decoders to be able to decode such encoded sequences.

Number of profiles which are provided by this standard is almost a couple of tens. In each of them, encoder is allowed to use a set of parameters and not allowed to use others depending on what is defined by the profile. Out of these profiles, at this section, we focus on those that are suitable for Scalable Video Coding (SVC) and those that are intended for 2D video compressions. For 3D applications, an extension is introduced named Multi-View Video Coding (MVC) which is not defined in this study.

1. Scalable Baseline Profile (SBP) is intended for applications in which the complexity of decoding must be low. Most of surveillance applications and mobile broadcast are in these group of applications.
2. Scalable Constrained Baseline Profile (SCBP) restrictions of SBP mostly still exist in this profile. However, the decoding complexity is a bit more than SBP and in real-time applications, this profile would be used.
3. Scalable High Profile (SHP) For streaming and storage purposes, this profile is an appropriate choice. Restrictions of SBP and SCBP do not exist here and coding with arbitrary resolution ratios is supported.
4. Scalable High Intra Profile (SHIP) is used where ever professional applications using H.264 are targeted.

### Levels

Levels make it possible to manage maximum data rate that can be served in an acceptable range of quality on the link. In other words, it is used for bandwidth utilization controlling purposes. More over, different levels provide options to control maximum resolution and solve possible memory issues which might happen at the decoder side. To clarify, lower levels mean that the maximum bit rate and resolution must be below or equal to a specific level. The idea of controlling max resolution is to make it possible for devices with different capabilities to be able to decode and play back H.264 encoded bitstream.

For example, assume that baseline profile is chosen for encoding input bitstream which imposes its encoding rules on the encoding parameters. Then, using levels, one can guarantee that the picture resolution would fit the targeted resolution and the bit rate and frame rate would be in a range that is acceptable for a specific device. It must be mentioned that, while a level is chosen, the decoder must be capable of handling all levels below thatl for the defined profile.



### 4.3 Trade-offs in H.264

This study investigates trade-offs on wireless networks for which multimedia is one of the most important applications. Transmission of voice over IP networks and its challenges and the trade-offs has been presented in Chapter 3. However, in case of video content, it is more complex to compare compression, QoE, and energy efficiency.

H.264 standard has been studied in this paper for which among all profile introduced in previous section, only baseline and constrained baseline profiles can be used for wireless multimedia communications. However, it is not possible to simply change between these profiles and calculate the bit rate and compression ratio for video codecs. The reason is that, the same profile at the same level of operation for two different video sequences might show totally unequal results.

On one hand, video slides have various motion conditions. On the other hand, encoding each video sequence highly depends on these movements of the video. The reason relates back to the coding schemes which use different coding techniques for motion compensation. As a result, encoding a video sequence containing 100 frames of a smooth nature scene requires much less encoding time and energy than coding a high movement sequence of 100 frames of a football match.

However, it is necessary to find an approach to compare three above-mentioned components on video codecs as well. The reason is that encoding video sequences require much more energy and time than voice data. More over, encoding video bitstreams result in high data rates and their transmission also impose a high power consumption cost. Choosing an inappropriate codec might result in either a very low battery lifetime duration which is not acceptable for users or an acceptable battery usage but poor quality.

In the following sections, at first the method used is described and then according to the methodology different components of tradeoff are compared to each other.

#### 4.3.1 Methodology

To analyze the effect of encoding parameters on bit rate, complexity, and power consumption in the context of video sequences, two issues must be considered and solved. Firstly, the parameters must be chosen in a way that encoding and decoding processes be applicable for real-time applications. For this reason, data rate of the output must be tenable for these applications. Complexity of coding algorithm is another issue which might be so that portable devices processors be able to do handle it. Power consumption must be also acceptable for batteries of these devices.

Since H.264 is a widely used standard these days, it is better to see tradeoffs on its codecs rather than other old algorithms. However, as it is mentioned previously

only baseline profile can be used and there is a need to see different levels of baseline profile effect on intended parameters. Another main problem is that it is not possible to use different sequences and encode them using available codecs and compare them with each other as motion of sequences has a great impact on the result. To solve this, the scope of this project is only on sequences with low/medium motion conditions.

To overcome the problem of profiles which can be used, baseline profile of H.264 can be divided into 9 different complexity levels. Each of these levels are called Complexity Parameters (CP) which is similar to what is done by [21]. Each complexity parameter corresponds to a certain amount of processing level that are separated by the coding algorithms used in them. These differences can be seen in the searching techniques they use, prediction schemes and other factors which can strictly affect both efficiency and complexity of encoder.

Using this approach, makes it possible to compare power consumption of different practical codecs. However, these CPs do not affect data rate, PSNR, and SSIM that are important factors. In other words, 9 different codecs which are all suitable for real-time applications and consume different amount of energy are in hand. However, the tradeoff between QoE and compression is not achieved yet.

Calculating encoding energy consumption for different complexity parameters of H.264 baseline profile, enable us to calculate PSNR and bitrates. For this purpose, this study takes advantage of different quantization steps named Quantization Parameter (QP). for different QP values and different sequences. The most important assumption that is used is that for a given QP value, the bitrate and PSNR will almost remain the same.

In other words, as it can be seen in [21] and [22], data rate for the CP values between 2 to 10 for a given QP do not change noticeably. The change in PSNR is less than 1 dB in total and the change in bit rate is almost 1 kbps. In addition to that, within a single CP changing QP does not effect cycles per second (MHz) of CPU clock which is used to calculate processing power.

This paper, uses these assumptions and scales different CPs from 1 to 9. In this range, both data rate and PSNR can be assumed constant. However, there are two things that affect data rate i.e. different QP values and also different sequences. The latter is because of the motion of different sequences. In this paper four different sequences with a medium to low motion are considered to minimize the impact of motion of sequences.

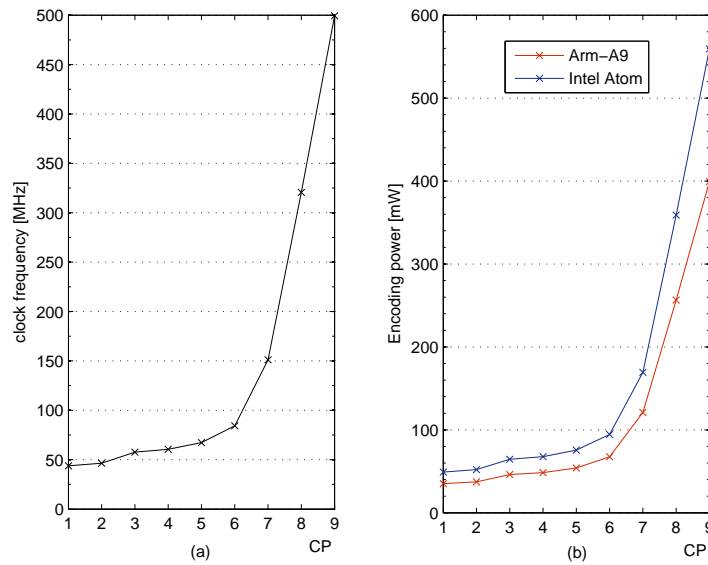


Figure 4.4: Different CPs within baseline profile frequency clock and encoding power.

### 4.3.2 Energy required for compression and transmission

#### Processing power

This subsection investigates different aspects of energy requirements using introduced method. At first, processing power of baseline profile for each of those CPs are calculated. For transmission power, the same algorithm used in audio codecs would be applied. I.e. having in hand data rate for different QP values of a specific CP makes it possible to estimate power consumption for transmission over different technologies.

Figure 4.4.a maps different complexity levels to their corresponding clock frequency in MHz. This figure presents a monotonic ascending schemes in clock frequencies of different levels of CP. Then using this frequencies, it is possible to calculate encoding power consumption for different processors.

To calculate encoding power first one must find an appropriate processor. In this study two newly released processors which are used in many smart phones and tablets have been investigated. Intel atom with x86-64 instruction set and ARM Cortex-A9 with instruction set ARMv7 are chosen for this purpose. It must be noted that they will be used in dual-core or quad-core mode on powerful devices.

According to [33] Intel dual-core processor consumes 1.12 mW per MHz and ARM Cortex-A9 consumes 0.8 mW. Knowing this amount of energy consumed per cycles per second and also Figure 4.4.a one can calculate processing power consumption for different CPs using these two processors. Figure 4.4.b presents this. As it can

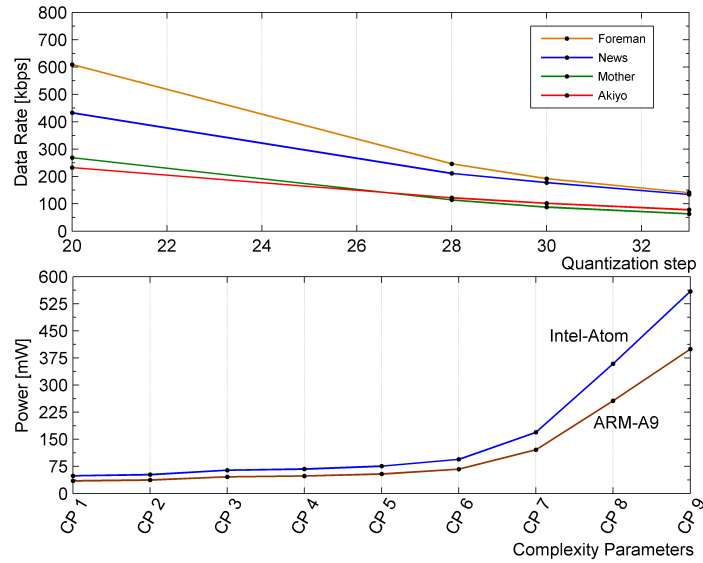


Figure 4.5: Power consumed in compression in different data rates.

be seen Intel processor consumes more energy than ARM.

After all, to see the efficiency of different CPs in the context of compression and processing power consumption, it is required to consider Figure 4.5. This figure presents the trade-off between data rate of the output of encoder and the power that encoder consumes to encode in that data rate. Both processors are again compared in this figure.

### Transmission

As it is already mentioned, the factor which affects transmission energy is QP as changing QP result in variation of data rates. The trend of transmission power change has a direct relation with QP. Increasing QP will reduce data rate which result in decreasing the amount of energy required for transmission.

To see the effect of QP on transmission power better, it is required to see its influence on different sequences. For this reason, Four different YUV sequences are investigated here. For each sequence the impact of 4 different QPs of 20, 28, 30, 33 are examined and the transmission power over different transmission technologies is calculated.

Table 4.1 presents calculated values of these parameters. As it can be seen in the four tables provided for different sequences, at QP=20 the highest amount of energy is required for transmission while QP=33 requires least amount of energy as expected. The reason relates back to the data rate. However, impact of higher QPs on QoE which would be examined in the following sections must not be neglected.

Another important factor which can be proven by Table 4.1 is the possibility of trading-off power consumption of encoding and transmission. This issue was

Table 4.1: Transmission energy for different sequences [mW].

(a) Foreman

QP	LTESISO	LTECLSM	WiMAX	UMTS	Edge	WiFi TTI=300
20	15835	5481.2	974.4	6090.2	522.7	101.5
28	6405	2217.1	394.1	2463.4	211.4	41
30	4986	1725.8	306.8	1917.6	164.6	31.9
33	3659	1266.7	225.2	14.07.4	120.8	23.4

(b) News

QP	LTESISO	LTECLSM	WiMAX	UMTS	Edge	WiFi TTI=300
20	11256	3896.2	692.6	4329.1	371.6	72.1
28	5489	1900	337.8	2111	181.2	35.2
30	4613	1596.7	283.9	1774	152.3	29.6
33	3482	1205	214.27	1339	114.9	22.3

(c) Mother and Daughter

QP	LTESISO	LTECLSM	WiMAX	UMTS	Edge	WiFi TTI=300
20	6986.2	2418.3	429.9	2687	230.6	44.8
28	2969	1027.7	182.7	1419	98	19
30	2290.1	792.7	140.9	880.8	75.6	14.7
33	1645.8	569.7	101.3	633	54.3	10.5

(d) Akiyo

QP	LTESISO	LTECLSM	WiMAX	UMTS	Edge	WiFi TTI=300
20	6042.9	2091.8	371.9	2324.2	199.5	38.7
28	3166.3	1096	194.9	127.9	104.5	20.3
30	2647.1	916.3	162.9	1018.1	87.4	17
33	2034	704.2	125.2	782.4	67.1	13

introduced in audio codecs as well. However, in that case, the energy for transmission was mostly more than processing energy as compression of voice is not that complex. This condition is totally different tconsidering video codecs. Most of the energy required for processing procedure of video codecs is in encoding phase.

### 4.3.3 PSNR and SSIM

QoE of video is evaluated in two general cases in this thesis. Once after encoding/decoding process and one more time after transmission to see the impact of transmission channel. In both of these two cases, two metrics have been investi-

Table 4.2: Bitrate, PSNR, and SSIM after compression of H.264 different complexity parameters.

Sequences	Parameters	20	28	30	33
<b>Foreman</b>	Bitrate	609.02	246.34	191.76	140.74
	PSNR	42.56	36.89	35.53	33.57
	SSIM	0.98	0.95	0.94	0.93
<b>News</b>	Bitrate	432.91	211.1	177.41	133.92
	PSNR	43.04	37.24	35.61	33.47
	SSIM	0.98	0.96	0.96	0.94
<b>Mother and Daughter</b>	Bitrate	268.7	114.19	88.08	63.3
	PSNR	43.45	37.98	36.63	34.78
	SSIM	0.98	0.95	0.91	0.91
<b>Akiyo</b>	Bitrate	232.42	121.78	101.81	78.24
	PSNR	44.38	38.78	37.3	35.25
	SSIM	0.98	0.97	0.96	0.94

gated to have a more precise result. PSNR and SSIM are those two measurement metrics which are considered.

### QoE after compression

In this scenario video sequences are encoded and decoded and finally the PSNR and SSIM are calculated. This process is repeated for all four sequences and QP values. JM v.15 which is the reference H.264 codec is chosen here. Parameters in the configuration file were changed for each complexity parameter accordingly.

Parameters which were required to be changed in each encoding run configuration were QP, searching algorithms, the format in which encoder created its output which could be either RTP or .264, profile of codec. In all encoding procedures the frame per second parameter was set to 30 and 102 frames were chosen from the YUV sequence to be encoded all in QCIF format.

Simulation results of encoding process for PSNR, and SSIM was as expected. Table 4.2 presents the result obtained from these simulations for all four sequences and QP values. The trend of decrement in PSNR, SSIM, and bitrate by increasing QP, is presented in the table. The highest SSIM obtained in all cases is equal to 0.98 out of 1 in the least QP equal to 20. As it is already mentioned, sequences motion are chosen to be in an acceptable range of similarity with each other. As a result, it is convenient to compare these values with each other. Bitrate is given in kbps and PSNR is in dB.

Although it is always better to consider all factors that affect QoE, calculations of PSNR and SSIM which are presented in Table 4.2 are important as they show the efficiency of encoder of the codec. Decoder side of the codec has some options

for concealing errors like packet losses which are not enabled in this specific part of experiment. The reason is that it is desired to see how a codec would operate when channel is ideal.

In this condition, original sequence would be compared to the decoder output and encoder reconstruction output. The latter is produced in encoding process and used while decoding steps as reference pictures. Number of references in all parts of this study is considered to be 1.

### **QoE after transmission**

PSNR and SSIM of QoE after compression is obtained without considering any defect in the wireless channel. However, it is not true in real applications. Different sources of error exist which will considerably decrease both of them and generally the QoE of video codec.

Among various error resources, packet loss error is investigated here. These packet losses are an inherent characteristic of wireless channels due to many reasons and in real life applications, users face them repeatedly while watching an online television or video-conferencing. Degradation range due to this issue might change between "not considerable" to "unacceptable".

This range not only varies as a function of Packet Loss Rate (PLR) but the pattern of packet losses affect video quality as well. For this reasons different factors in various conditions must be examined ensuring that different aspects of QoE degradation because of PLR and its pattern is covered.

The best way to see effect of packet losses in different rates on the QoE is transmitting encoded sequence over channel and then seeing what happens. However, this method is time consuming and requires some additional devices. In order to see the effect of PLR, this study uses MATLAB to simulate channels and drop packets.

The first thing that must be done is to encode YUV sequence in RTP mode using JM. According to the number of frames which are encoded, a random sequence of ones and zeros is generated. These numbers are then used to drop or keep a certain packet. It must be noted that in this approach the first three packets of RTP encoded sequence must be kept unchanged or decoding process will fail.

The next step is to find a way to drop or keep encoded packets for which the above-mentioned random sequence of numbers would be used. Each zero in that sequence means that the packet must be kept and ones imply dropping packets. In order to have a better result in this approach, it is required to encode at least 100 frames.

Then, it is possible to see the effect of packet loss for different rates by changing ratio of ones to the whole number of numbers in the sequence. Moreover, it is also possible to change the pattern of packet losses to see also the effect of consecutive

drops. This study considers these drops random but it is possible to define a certain pattern.

After encoding and dropping packets it is time to decode the YUV bitstream. Here the algorithm which H.264 must use for concealing packet losses occurred is selected. Frame copy and motion copy are available algorithms in JM. During decoding process it has been seen that for PLRs more than 12 percent, decoder could not operate correctly.

Table 4.3: PSNR and SSIM under different PLRs and QP cases shown in  $C_{PLR}^{QP}$ .

(a) QP=20,28

Seq.	Parameters	$C_0^{20}$	$C_3^{20}$	$C_5^{20}$	$C_{10}^{20}$	$C_0^{28}$	$C_3^{28}$	$C_5^{28}$	$C_{10}^{28}$
Foreman	PSNR	42.7	29.67	28.77	27.57	36.69	29.39	28.47	26.66
	SSIM	0.98	0.89	0.78	0.75	0.95	0.9	0.79	0.61
News	PSNR	43.04	37.39	29.73	30.03	37.24	33.8	37.07	28.95
	SSIM	0.98	0.9	0.81	0.72	0.96	0.89	0.79	0.72
Mother and Daughter	PSNR	43.57	37.64	36.44	36.91	37.98	35.24	34.45	34.53
	SSIM	0.98	0.87	0.81	0.75	0.95	0.9	0.86	0.8
Akiyo	PSNR	44.38	40.99	37.47	37.3	38.78	36.91	34.89	35.03
	SSIM	0.98	0.94	0.85	0.82	0.97	0.91	0.83	0.79

(b) QP=30,33

Seq.	Parameters	$C_0^{30}$	$C_3^{30}$	$C_5^{30}$	$C_{10}^{30}$	$C_0^{33}$	$C_3^{33}$	$C_5^{33}$	$C_{10}^{33}$
Foreman	PSNR	35.18	29.19	28.17	26.26	33.17	28.58	27.7	25.89
	SSIM	0.94	0.89	0.78	0.62	0.93	0.87	0.78	0.6
News	PSNR	35.61	32.61	28.72	28.59	33.47	31.47	29.07	28.29
	SSIM	0.96	0.92	0.82	0.79	0.94	0.89	0.78	0.71
Mother and Daughter	PSNR	36.63	34.42	33.7	34.01	34.78	32.94	32.46	32.43
	SSIM	0.91	0.89	0.8	0.76	0.91	0.88	0.85	0.76
Akiyo	PSNR	37.3	35.67	34.12	34.0	35.25	33.92	33.11	32.75
	SSIM	0.96	0.91	0.85	0.79	0.94	0.87	0.79	0.78

Table 4.3 summarizes PSNR and SSIM values of different sequences under packet loss conditions. Expectedly, general trend presents decrement in both PSNR and SSIM for all the sequences. All of these calculations are obtained while IDR is set to zero in encoding procedure. Changing this parameter to 10 would result in a considerable improvement in the quality of output. Simultaneously, rate will increase as well.

To see the effect of this parameter, PSNR of foreman sequence under 4 different QPs in two cases have been calculated. PSNR without error and with 10 percent packet loss rate. To make the impact of pattern of losses, the latter case has been examined 5 times and the average of PSNR over these times are taken into account.



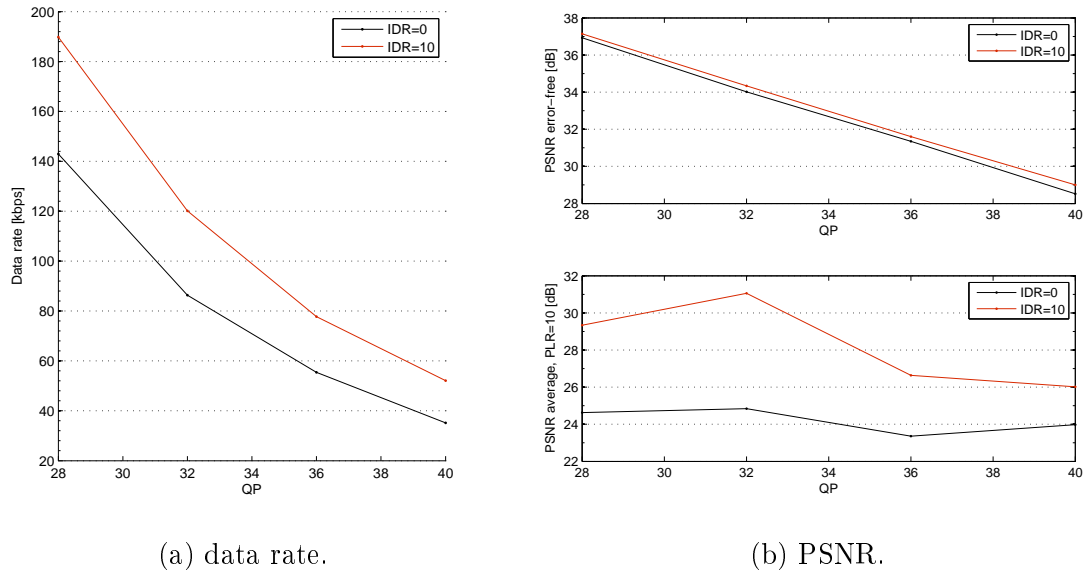


Figure 4.6: Impact of IDR on data rate and PSNR.

Figure 4.6 presents the impact of IDR on data rate and PSNR over all these different conditions. As it is shown in the figure, trade-off between data rate and PSNR for IDR disabled and enabled cases in error-free condition might not be reasonable. I.e. the amount of increment in data rate is more than improvement in quality of output.

When there are packet losses it might be tenable to use IDR. In fact, it enhances motion compensation and as a result when they are used as reference frames a better prediction of blocks would be achieved improving PSNR considerably. This effect is obviously shown in Figure 4.6.b where in some cases PSNR would be close to the error-free corresponding one which means compressed video is more faithful to its original.

In addition to PSNR improvement, complexity of codec would be higher when IDR is used. To conclude, trading off all these parameters would be even more complex in this scenario. A noticeable better quality in cost of more transmission power consumption, higher bandwidth requirement, and more processing power can be obtained.

## 5. CONCLUSION

This thesis examines the trade-offs between quality of experience, compression, and energy consumption. Energy consumption is divided into energy required for transmission over different technologies and energy spent for processing compression algorithms in voice and video codecs.

Results obtained in the thesis prove that implementing a system that modifies factors involved in media streaming under different conditions provides the best possible performance in multimedia real-time applications. For example, changing data rate based on network condition, would improve QoE. Another example is seen in video codecs, where modifying QP makes it possible to trade-off between data rate, QoE, and even energy consumed for transmission.

In addition to the above-mentioned parameters, the compression algorithm used in different voice codecs and in different levels of baseline profile of H.264 video codec, affect both QoE after transmission and power consumption. More complex algorithms increase the compression ratio, which conserves transmission energy. However, packetizing more and more bits and bundling them into one UDP frame (which is the most used protocol in real-time applications) under poor network conditions with high packet loss ratio degrades QoE severely; each lost packet means huge amounts of lost data.

Introduced trade-offs in the project are not constant. For example, new air interface technologies bring different energy consumption values; these affect total calculations of energy requirements, which must be considered in the decision-making process. Improvements in network conditions also make it possible to choose more complex codecs with higher compression ratios without any QoE degradation. The same possible changes in trade-offs happen by introducing new compression techniques.

The possibility of changes in trade-offs, however, does not mean that implementing the system is impossible. Considering all potential changes, such a system can still be implemented and be valuable for certain parameters of multimedia streaming applications in portable devices. These systems provide users with high performing, long-lasting devices.

Comparing energy consumed for transmission and compression, which is an important metric in trade-offs, affect other parameters. This clarifies one reason why

implementation of such a system is important. For example, the coding process of foreman sequence in QCIF resolution varies from 35 mW in the lowest complexity mode in a ARM-A9 dual core processor to 559 mW at the complexity level 9 of baseline profile in an Intel Atom dual core processor. Transmission power for the same sequence and in the same resolution varies between 23 mW while the QP is set to 33, resulting in the lowest PSNR considered in this paper, to 15835 mW over LTE SISO setting QP to 20, yielding the best PSNR.

These values for transmission and compression energy consumption proves the importance of minimizing energy consumption while obtaining the best possible quality. However, this thesis considers QoE as a single variable estimated by objective measurement methods, that is just a basic assumption of quality of experience. In order to have a more accurate estimation of QoE, different factors, such as the context of application, the user's characteristics, and more technical influence factors such as the protocol used for transmission, must be considered.

Any change in one factor might improve or degrade another factor, affecting the final QoE parameter. On the other hand, at each instant of time  $t_i$  for a specific application, there exists a set of factors that might be replaced with other ones at time  $t_{i+1}$ . There could be some cases in which these factors are not replaced, but their influence weight could change. This weight could be the direct effect of a single factor on QoE or the effect of its correlation with some other components.

Considering the potential influence of these metrics on QoE, and examining trade-offs between these components is quite complicated without performing a general but precise study. This paper tried to prepare the background for such a future research project while maintaining accuracy and examining various parameters affecting the trade-offs.

Putting all of these parameters together and viewing the results in one figure would be helpful. However, because of the variety of parameters, this kind of figure might be confusing, specifically in the case of video codecs. Figure 5.1 presents all of these trade-offs covering different audio codecs.

In this figure, the QoE estimator is considered the well-known MOS value. The minimum level of acceptance for QoE, according to MOS grading values, is 3.5. Different codecs are replaced with numbers in the x-axis, as the most important feature of each codec in this context is its complexity. The y-axis is divided into two parts. The first part presents values of one to five for QoE, and the second part, which varies between zero to six, presents energy consumption for each complexity level in mW. The DSP considered for encoding is Cortex M3, and the technology over which encoded packets are transmitted is WiFi with TTI value below 15 ms.

According to Figure 5.1, for some complexity levels, which are nominators of one or several different audio codecs, obtaining better QoE while maintaining or reducing

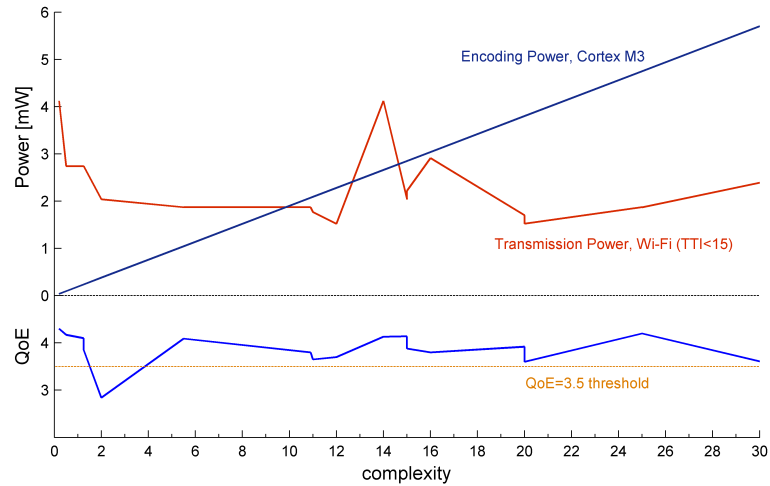


Figure 5.1: Trading-off QoE, compression ratio, and energy efficiency in audio codecs.

power consumption might not be feasible. In this context, one might consider the total power as the sum of encoding and transmission powers. At the complexity levels 12 and 14, there is a slight increase in QoE, while the total power consumption shows a noticeable increase. Another example is the codec with complexity level 2; here, the total energy consumption is very low, but the QoE of this codec would not be tolerated in certain real time applications such as VoIP.

An important point raised by this figure is the ability of codecs to operate despite impairment. All calculations were done assuming perfect network conditions (in contrast to video codecs in which different packet loss rates have been studied during the calculation process of PSNR and SSIM). This is one reason why some codecs with higher complexity levels result in lower MOS values compared to some less complex codecs. If the same calculations were performed considering different values of delay, jitter, and packet losses the curve would have revealed different results.

## BIBLIOGRAPHY

- [1] G. Bernardi, P. Buneman, and M. K. Marina. Tegola Tiered Mesh Network Testbed in Rural Scotland. In WiNS-DR 08, San Francisco, CA, Sept. 2008
- [2] V. Pejovic and E. Belding. Energy-Efcient Communication in Next Generation Rural-Area Wireless Networks. In CoRoNet10, Chicago,IL, September 2010
- [3] ITU-T, "Batteries for portable ICT devices," ITU-T TechWatch Alert, Feb. 2010
- [4] R. Trestian, A-N. Moldovan, O. Ormond, and G-M. Muntean, Energy Consumption Analysis of Video Streaming to Android Mobile Devices, IEEE/IFIP Network Operations and Management Symposium, (NOMS), April 2012
- [5] M.zorzi, R. R. Rao, "Error control and energy consumption in communications for nomadic computing," IEEE transactions on computers, March 1997
- [6] R. Trestian, O. Ormond and G-M. Muntean, "On the Impact of Wireless Traffic Location and Access Technology on the Mobile Device Energy Consumption", The 37th IEEE Conference on Local Computer Networks (LCN), October 2012
- [7] T. Chen, H. Kim and Y. Yang, Energy Efficiency Metrics for Green Wireless Communications, International Conference on Wireless Communications and Signal Processing (WCSP), pp. 1–6, 2010.
- [8] Codecs, A reference guide to all things VOIP, from <http://www.voip-info.org>, June 17, 2012
- [9] A.D. Clark, Modeling the effects of burst packet loss and recency on subjective voice quality, 2nd IP-Telephony Workshop, Columbia University, New York, April 2001.
- [10] Rosenbluth, J. H., Testing the Quality of Connections having Time Varying Impairments. Committee contribution T1A1.7/98-031
- [11] Alain Hore, D. Ziou "Image quality metrics: PSNR vs. SSIM," IEEE International Conference on Pattern Recognition, 2010
- [12] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, Image quality assessment: From error measurement to structural similarity, IEEE Trans. Image Processing, vol. 13, Jan. 2004.

- [13] Z.wang, E. P. Simoncelli, and A. C. Bovik, "Multi-Scale Structural Similarity For Image Quality Assessment," 37th IEEE Asilomar Conference on Signals, Systems and Computers, Nov. 2003.
- [14] B. Pham, O. Wong, "Handheld Devices for Applications Using Dynamic Multimedia Data," Proc. 2nd International Conference on Computer Graphics and Interactive Techniques in Australasia and South East Asia, Singapore, 123-130, 2004.
- [15] BJORK, S., BRETAN, I., DANIELSSON, R., KARLGREN, J. AND FRANZEN, K. (1999). "WEST: A Web Browser for Small Terminals." CHI Letters Vol 1, 1: 187 - 196.
- [16] LI, C.-S., MOHAN, R. AND SMITH Multimedia content description in the InfoPyramid. Proceedings of the 1998 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1998
- [17] E.Baalkini, H.Sbeiti, S.Niar, "H.264 Macroblock Line Level Parallel Video Decoding on Embedded Multicore Processors," Digital System Design (DSD), 2012 15th Euromicro Conference on 5-8 Sept. 2012
- [18] J-Y. Chen, C-W. Chiu, G-L. Li, M-J. Chen "Burst-Aware Dynamic Rate Control for H.264/AVC Video Streaming," Broadcasting, IEEE Transactions on March 2011
- [19] X. Chun-Dong, L. Yi-Qing, Z. Mu-Gui, "A Design of Embedded Video Server Based on H.264," Communication Systems and Network Technologies (CSNT), 2012 International Conference on 11-13 May 2012
- [20] E.S.C.Takahashi, Application aware scheduling for power management on IEEE 802.11, Proc. IPCCC., Feb. 2000,
- [21] W. Pu, Y. Lu, F. Wu, "Joint Power-Distortion Optimization on Devices with MPEG-4 AVC/H.264 Codec," International Conference on Communications (ICC), vol. 1, pp 441-446, 2006
- [22] Z. Chen, P. Zhou, Y. He, "Fast Integer Pel and Fractional Pel Motion Estimation for JVT," Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG, Awaji, Island, JP, 5-13 December, 2002
- [23] Conti, J.P. , "Talk about a change [VoIP replacing PSTN," IET JOURNALS and MAGAZINES, Volume: 50, Issue: 12, 2004

- [24] Z. Stojanovic, D. Babic, "Traffic Engineering for VoIP Network based on PSTN Statistical Models," 9th International Conferences on Telecommunication in Modern Satellite, Cable, and Broadcasting Services (TELSIK09), 564-567, October 2009
- [25] Y. Zhang, SIP-based VoIP network and its interworking with the PSTN, *Electron. Commun. Eng. J.*, vol. 14, no. 6, pp. 273-282, Dec. 2002.
- [26] B. Ngamwongwattana, A. Sombun "Enhancing Adaptive Control in VoIP by Characterizing Congestive Packet Loss," 11th International Conference on Advanced Communication Technology, 2009.
- [27] L. Caviglione, "A Simple Neural Framework for Bandwidth Reservation of VoIP Communications in Cost-Effective Devices," *IEEE Transactions on Consumer Electronics*, Vol. 56, No. 3, August 2010
- [28] P. C. K. Hung, M. V. Martin "Security issues in VoIP applications," Canadian Conference on Electrical and Computer Engineering, 2006. CCECE '06.
- [29] D. Butcher, X. Li, J. Guo "Security Challenge and Defense in VoIP Infrastructures," *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETIC-SPART C: APPLICATIONS AND REVIEWS*, VOL. 37, NO. 6, NOVEMBER 2007
- [30] Salami, R. and Laflamme, C. and Adoul, J.-P. and Kataoka, A. and Hayashi, S. and Moriya, T. and Lamblin, C. and Massaloux, D. and Proust, S. and Kroon, P. and Shoham, Y. "Design and description of CS-ACELP: a toll quality 8 kb/s speech coder," *Speech and Audio Processing*, IEEE Transactions on 1998
- [31] A. Clark, "Modeling the Effects of Burst Packet Loss and Recency on Subjective Voice Quality," 2nd IPTel Workshop, 2001, pp.123-127.
- [32] [www.ti.com](http://www.ti.com), Texas Instruments company
- [33] Linley Gwennap, "BCM 5000 Provides Dual Issue, Multithreading for Broadband Chips," The Linley group, [www.MPRonline.com](http://www.MPRonline.com), 2010